UNIVERSITY OF EASTERN FINLAND
PHILOSOPHICAL FACULTY
SCHOOL OF HUMANITIES
FOREIGN LANGUAGES AND TRANSLATION STUDIES
English language and translation

Alma Inkeri Tuominen

INVESTIGATING USABILITY: A CASE STUDY OF WORDFAST PROFESSIONAL

Pro Gradu Thesis
May 2012

ITÄ-SUOMEN YLIOPISTO - UNIVERSITY OF EASTERN FINLAND

Tiedekunta – Faculty **Philosophical Faculty**	**Osasto – School** **School of Humanities**
Tekijät – Author **Tuominen Alma Inkeri**	
Työn nimi – Title Investigating usability – a case study of Wordfast Professional	

Pääaine – Main subject	**Työn laji – Level**		**Päivämäärä – Date**	**Sivumäärä – Number of pages**
English language and translation	Pro gradu -tutkielma	**x**	**25.5.2012**	**86**
	Sivuainetutkielma			
	Kandidaatin tutkielma			
	Aineopintojen tutkielma			

Tiivistelmä – Abstract

The aim of this study was to investigate the usability of Wordfast Professional from the point of view of novice translators to form a better understanding of what causes usability issues for translators. This was achieved by organising a usability test. Little research has been conducted about usability in the translation industry, and it seems usability has often been a secondary goal in designing electronic translation tools.

This study began with a review of relevant literature in both translation studies and usability research. The basics of usability testing were introduced, and based on those the plans for the usability test in this study were drawn. In the usability test, six translation students were asked to translate an excerpt from a user manual using Wordfast Pro, a translation memory tool. After translating the participants answered a questionnaire, which included questions about how easy it was to use the software and what the participants thought of it. Each test was also recorded using a screen recording software to provide further assistance in analysing the process.

All of the participants were students of English language and translation at the University of Easter Finland, and had background knowledge of translation memories as they had all attended at least one course in translation technology. Some of them had work experience in translation, but that did not seem to directly affect the results in all of the tasks.

In the usability test, the participants were asked to create a new project and a translation memory, translate the text, and clean the file at the end. The participants completed the given tasks with varying levels of success and satisfaction. Based on their answers in the questionnaires and the data from the screen recordings, some usability issues were identified. Some features of the software caused the participants more confusion than others. The most severe usability issues were related to creating a new project and starting a new translation. On Wordfast, the user has to create a new project and a new translation memory separately, and this caused confusion when participants thought they only needed to create a new project and then start translating. Another feature that the participants had problems with was the file clean-up function, which was difficult to identify and find. Related to this, many participants thought that the symbols and icons on Wordfast were confusing and the menus difficult to navigate. The help section had some shortages and was also difficult to navigate.

Overall, the issues faced by the participants can be partly attributed to inexperience, but at the same time, the software should be self-explanatory enough for users who are familiar with the basics of translation memory tools. There may be other similarly inexperienced users who would like to buy the tool to use in their work. From the point of view of novice translators, easy learnability and memorability are important. This study has shown that Wordfast Pro is not a user-friendly tool in those aspects. However, more research with more participants is needed for statistically valid results.

Participants were also asked for their opinion on the teaching of translation technology at the university. On this, all of the participants were united: they all thought that more courses in translation technology should be offered, and that translation tools should be used on as many courses as possible. They did not think they had received the skills necessary in working life.

The results of this study are only indicative and more research in the particular field of translation is needed. Hopefully this study can act as a starting point for future testing.

Avainsanat – Keywords
Translation technology, translation tools, translation memory, usability, usability testing

TABLE OF CONTENTS

1.Introduction....1
2. Translation technology....2
2.1 Machine translation....2
2.2 Translation tools....5
2.2.1 Translation memory systems....6
2.2.2 Terminology tools....11
2.2.3 Corpora....11
2.3 Localisation....15
2.4 Professional translators and electronic translation tools....17
3. Usability research....19
3.1 The definition of usability....20
3.2 Usability engineering....22
3.3 Usability research in the translation industry....26
3.4 Questionnaires....30
3.5 Logging....33
3.6 Usability testing....33
3.7 Conducting usability research....37
3.7.1 Planning the usability test....37
3.7.2 Selecting participants and user characterisation....38
3.7.3 Other materials needed for the test....40
3.7.4 The test situation....42
3.7.5 Evaluation....44
4. Research material and method....45
4.1. Wordfast Professional....48
5. Analysis and discussion....49
5.1 Test results – questionnaires....50
5.1.1 Appearance and layout....50
5.1.2 Functional issues....51
5.1.3 Issues with online help....54
5.1.4 Positive feedback....54
5.2 Test results - screen recordings....55
5.2.1 Setting up....55
5.2.2 Translation....57
5.2.3 Saving and cleaning....60
5.2.4 Online manual....62
5.3 Overview and categorisation of usability issues....63
5.4 Translation technology training....66
5.5 Methodological evaluation....67
6. Conclusion....69
REFERENCES....72
FINNISH ABSTRACT
APPENDICES

1. Introduction

Usability is an important aspect of software development and considerable effort is often put into making user interfaces easy to use. Usability includes notions such as how easy it is to start using a software, how easy it is to accomplish tasks with it, how quickly users perform tasks with it, how many errors users make, and how pleasant it is to use (Nielsen 1993:25). If an interface is difficult to use, users most likely will not want to return to it and are more likely to choose another tool. This does not appear to be the case in the translation industry, however: there is only a limited number of suitable tools on the market which limits the choice; in addition employers often dictate which software they want their translators to use.

The aim of this study is to investigate the usability of Wordfast, a translation memory software, and uncover possible problem areas. Translation aid software, or electronic translation tools, are an integral part of modern translators' equipment, but the usability of different translation tools is often thought to be rather poor and the programs difficult to use. The objective in this study is to acquire a better understanding of what poses problems for translators when they use the translation tool in question.

Usability testing is an aspect of electronic translation tool development that has been little researched. There are few sources that concentrate particularly on translation tools - sources on usability testing in general are abundant however. The parties who develop electronic translation tools seem to be situated quite far from the pool of translators who use their products in their work in terms of contact and customer feedback (Jääskeläinen and Mauranen 2001: 359-360). The underlying assumption for this study is that translation tools and translation memories in particular are difficult to use in translators' opinion and that improvements are wished for. This assumption is based mainly on personal experience. This study sets out to explore whether novice translators, i.e. students who study translation at the University of Eastern Finland, will experience the same. This is accomplished through a usability test in which a group of participants

carry out a test translation using Wordfast Pro and then complete a survey. Screen recordings are used to further investigate the test process. The results are then analysed in light of the usability research basics.

In chapter 2, the context for this study is described in translation studies and in chapter 3, in the field of usability research. A brief overview of machine translation and translation tools is given in 2.2. Then, in section 3, the theoretical background in usability research is explained and the basics of usability research and usability testing are presented. Next, in chapter 4, the research material and methods are presented. Chapter 5 includes the analysis of the results and discussion. Finally, chapter 6 concludes the study.

2. Translation technology

2.1 Machine translation

According to Somers (2003: 4), the first occurrences of computers used in translation date back to the 1960's, and virtually the beginning of the computer age. Since then, computers have, in some form or another, been used to help with translation, even if not on a commercial or a wide-spread basis to begin with. However, it was only the Second World War that broke ground: in the war computers were used for code-breaking, and it was thought that translation could be performed in a similar way (Somers 2003:5). This was the beginning of machine translation, commonly known as MT. In the USA, MT was of special interest to the Massachusetts Institute of Technology, who ordered a large-scale research project on it. In 1964 a committee called Automated Language Processing Advisor Committee (ALPAC) was appointed to evaluate the profitability of MT research. The results were not encouraging, and in result machine translation research quickly declined all over the world. It is not surprising: technology at the time was primitive, and it shows the full difficulty of machine translation that to this day, fully functional, sophisticated, high quality machine translators do not exist. On the other hand, the final report of ALPAC suggests that more funding should be perhaps

directed towards machine-*aided* translation research, and this was to become the future trend (Somers 2003:6).

In the 1970's and 80's MT developed more extensively in areas where the need for translation arose, for example in Canada, Japan and western Europe; on bilingual areas or countries in close contact with other countries and languages. It was soon recognised that fully automatic, high quality translation of unrestricted texts (FAHQT) was not a goal easily reached. The ideas of *sublanguage* and *controlled language* emerged, meaning that language should be somehow restricted or not complete, this is to say simple, to be able to be translated by a machine. In the 1990´s global communication started to become easier with the spreading of personal computers and the internet, and technological advances further widened the research in machine translation. As technology developed, so did the translation tools: computer-aided translation (CAT) was born (Somers 2003:6). Computer-aided translation tools appeared on the market. A new need for translations arose on the internet, which also created a new and an important medium for translators themselves. Machine translation evolved from primitive through PC-systems to online systems (Quah 2006: 65-66). Online machine translation systems are widely available today, and typing "online translator" into a search engine produces several million hits. Most online translators are for general purposes, and therefore the quality is not usually very high, but for the end-user who needs occasional translations and for whom only indicative results are adequate, these systems are a great advantage even when the translations only form a guideline of sorts of the source text contents (Quah 2006: 86).

As defined by the European Association for Machine Translation (http://www.eamt.org/mt.php), "machine translation (MT) is the application of computers to the task of translating texts from one natural language to another." In recent developments it has become clear that with current technology fully automatic machine translation is not possible, but human intervention is nearly always needed, mostly in pre- or post-editing. It is now widely accepted that MT systems produce only indicative results that are then edited by the translator; sophisticated results are not

expected, and in fact some have argued that the goal of research should not be on FAHQMT (Quah 2006:60). This brings forth the concepts of human-assisted machine translation (HAMT) and machine-assisted human translation (MAHT), the difference of which is not always obvious.

A distinction is made between different types of translation, according to the way in which computers and human translators are involved and who is primarily responsible for the translation (Bowker 2002: 4). MT stands for machine translators, where no humans are involved, and HAMT stands for human-assisted machine translation, where machines translate but human translators edit the text. The opposite, in a way, is MAHT, machine-assisted human translation, which means that humans translate with the help of computer tools. HAMT and MAHT together are also, and more commonly, referred to as computer-aided translation (CAT). In short, MT systems aim to replace the translator, whereas CAT systems aim to equip the translator with such software that increases productivity and facilitates translation (Bowker 2002: 4). This study focuses on CAT and more specifically machine-assisted human translation in form of translation memories.

Traditionally, many professional translators seem to have taken a strong aversion to machine translation because of the fear that machines would completely replace them. That fear, however, has turned out to be ungrounded. Current machine translation systems cannot replace the translator in that machines do not fully "understand" language and its nuances, even if machines translate faster than humans. Machine translation systems require a text to be fairly simple to be able to produce convincing translations, and, as Austermühl (2001: 1) argues: "…since MT systems neglect the communicative, cultural and encyclopaedic dimensions of translation, it is questionable whether they really provide 'translation' at all." When excellent results are not required and only a basic understanding of a text is needed, machine translation systems are convenient. Machine translators can be used to speed the human translation process and to polish translations, therefore professional translators can use them to their advantage (Quah 2006: 90). Arnold et al. observe that

> *Seeing MT as a threat to translator jobs is also pointless. The quality of translations that is currently possible with MT is one reason why it is wrong to think of MT systems as dehumanizing monsters which will eliminate human translators or enslave them. It will not eliminate them, simply because the volume of translations to be performed is so huge, and constantly growing, and because of the limitations of current and foreseeable MT systems.*
> (Arnold et al. 1994:8, quoted in Austermühl 2001: 154)

Considerable time has, however, passed since the above statement and technology has taken leaps in development. The systems today may correspond better with the needs of the user.

2.2. Translation tools

There are various electronic tools translators can utilise. They range from monolingual dictionaries and thesauruses to complicated translation memory systems. Dictionaries today are most often electronic, and searching for equivalents is easier and quicker than using traditional books. The same applies to all electronic databases and glossaries, which are also easier to keep up to date in electronic format. Electronic dictionaries can be monolingual, bilingual or multilingual, and the search functions are versatile because the user can use different search strings. Dictionaries are naturally of great importance to translators, but not by any means the only way the translation process can be facilitated. The most commonly used translation tools are translation memories, terminology management tools and corpora. Translation tools are often combined into a workbench, which Quah (2006: 94) defines as follows:

> *"A 'workbench' or a 'workstation' is a single integrated system that is made up of a number of translation tools and resources such as a translation memory, an alignment tool, a tag filter, electronic dictionaries, terminology databases, a*

terminology management system and spell and grammar-checkers."

These workstations are sometimes called Translation Environment Tools as well (Bowker and Fisher 2010:61). Although first mentions of translators' work stations or workbenches date back to the 1970's and 1980's, it wasn't until the 1990's and the prevalence of personal computers that they became more widely used. Electronic tools have since become a natural part of a translators' tool kit as it is safe to assume that most translators today use some electronic tools in their work (Dragsted 2005:85). The following sub-chapters will present the most common translation tools in more detail to give an overview of the translation technology context that this study is a related to, particularly translation memory tools.

2.2.1 Translation memory systems

According to EAGLES (Expert Advisory Group on Language Engineering Standards, http://www.ilc.cnr.it/EAGLES/home.html) a translation memory is "a multilingual text archive containing segmented, aligned, parsed and classified multilingual texts allowing storage and retrieval of aligned multilingual text segments against various search conditions." Translation memory systems were created to increase translation efficiency and productivity by automating the linguistic transfer from source to target text (Austermühl 2001: 134). Translation memory systems are translation tools that most often work in conjunction with a word processor. In principle, the software records the translated text and forms an aligned corpus (see 2.2.3 for more about corpora) of the source text and its translation. The text is divided into segments usually by sentence, and these segments (translation units) are recorded along with their translation. When translating, the software compares the source text against the previously translated entries, and if there is a matching segment it offers the pre-translated segment as an equivalent. The translator then has the choice to incorporate the suggestion into the translation or discard it (Austermühl 2001:134).

The fundamental idea is that a translation memory system (TM system) allows the translator to reuse and profit from the previous translations which would otherwise be "wasted" and not utilised, or using them would be too arduous or time-consuming. Translation memories can be used with any texts, but they work especially well with texts that contain a lot of repetition, and are of great importance to translators of special field texts, such as technical documents (Quah 2006: 93, Austermühl 2001: 139). Also, the longer the document, the more benefits a translation memory will normally bring to the user. TM systems only became commercially available in the 1990's, but are now an integral part of a translator's tool kit (Bowker 2002: 93).

Some benefits of using a TM system are the increase in income through increased productivity, elimination of repetitive translation tasks and consistency in terminology and formulation, but the problem with the latter is that possible mistranslations and errors are also reused and repeated, if not spotted and corrected (Austermühl 2001: 140, Dragsted 2004:93). It is also a concern that the translator may choose to reuse a previously translated segment which is a perfect match for the current segment, but fails to take into consideration the context from which the different segments come from. This could result in inappropriate reuse of segments (Dragsted 2004:93). Also, some repeat clients demand discounted translation prices for re-using previous, similar translations completed for them, which may eliminate the increase in income obtained through translation memories. Issues of ownership and copyright have also been debated (Kenny 2011:471).

Segmentation of text is not always straightforward. Most commonly it is done on sentence level because sentences seem to form a natural linguistic unit. That can, however, cause problems if the software does not recognise different punctuation marks. The software neither analyses the text, and therefore abbreviations and embedded sentences can result in nonsensical segments. Another issue is that source text segments may not directly correspond with the translated segments as different languages form sentences in different ways, and translators may wish to divide a source text sentence into several target text sentences, or vice versa (Bowker 2002: 95). Dragsted (2004:91)

suggests that segmentation on sentence level may force the translator to focus on the sentence more than would be natural, and hence to be more inclined to preserve the sentence structure of the source text than when compared with natural circumstances.

A TM system normally offers various kinds of matches for source text segments: the most common are exact, fuzzy and term matches. An exact or perfect match is, according to its name, an identical match both in terms of linguistics and formatting. This means that the sentences are linguistically alike, and have exactly the same punctuation, inflections etc. including formatting (e.g. italics, bold…). It is in the end up to the translator to decide whether to use the segment offered by the translation memory or not. The translator must also consider other factors, such as context and the client's preferred terminology (Bowker 2002: 96).

Fuzzy matches are matches that have something in common with the source text but are not exact matches. They are similar but not identical to the source segment. The user normally can adjust the amount of fuzziness from 1 to 99 per cent, where 99 produces the most accurate matches and 1 is least likely to produce similar matches. The problem is that if the fuzziness threshold is set too high, there is a possibility that the system will produce "silence": that is, it does not retrieve potential partial matches because they do not have enough similarities to the source segment, but in reality could be helpful in translating. At the other end of the scale, if the fuzziness is set too low, it can easily produce "noise": the retrieved matches will be too different from the source-text segment to have any value in the translating process. Many translators prefer the fuzziness sensitivity threshold to be somewhere between 60 and 70 (Bowker 2002: 99).

TM system developers are attempting to include new kinds of matching algorithms into the software, namely full matches and sub-segment matches. Full matches are in essence exact matches, but differ from them in that they do not differentiate between the so-called variable elements, i.e. numbers, dates, times, currencies, measurements and sometimes also proper names. Sub-segment matching means that matches are retrieved for small chunks of text, so even if a full segment does not match, a match can be found

for a small chunk of it; hence the name sub-segment. If classified, sub-segment matching falls somewhere between fuzzy and term matching. A cross-system between full and sub-segment matching is also being developed: it would first try to find exact or fuzzy matches for full segments, and if no matches were found, it would use increasingly smaller chunks to find matches (Bowker 2002:98-105). Dragsted found out in her study (2004:279) that sub-segment matching did increase the output from the translation memory, but that translators found it difficult to achieve a smooth translation process with such short segments. She suggests that TM systems should be developed to work on a wider segmentation method, but still take into consideration smaller units, i.e. sub-segment matching.

Many of the current matching algorithms have some limitations: they cannot handle inflection or derivation. For example, a TM system would not recognise “dishes” as the same word stem as “dish”, or “talked” as “talk”. This seriously limits the matches that are retrieved, considering translators could find relevant information and just render the inflection to match their current translation (Bowker 2002:106). That would be an important improvement in the functionality of translation memories for translators of languages such as Finnish, which inflects the forms of a majority of words depending on their role in the sentence.

Most TM systems work in conjunction with an associated terminology management software, and using the software the translator can build a bilingual term base, a kind of a dictionary. The TM system then automatically searches for words in the term base if no exact or fuzzy matches are found and proposes term matches. This means that the translator can retrieve some relevant information and translation equivalents even with no matches from the corpus of translations (Bowker 2002: 101).

As mentioned, a TM system is in essence a database of text. Information can be stored and retrieved. When a translator first starts using a TM system, the database is empty until the user starts storing data on it. The stored source and target texts form the database (Dragsted 2004: 88). Most translators find it useful to build smaller subject-

field databases instead of a large one covering all fields and subjects. Large databases easily produce bad matches or no matches at all, whereas smaller ones produce better results because not many false hits occur. Translations can be entered into a database through interactive translation or post-translation alignment, or sometimes with a combination of both (Bowker 2002:108).

Translation segments or units can be stored in a translation memory in two different ways. Interactive translation means that the translator translates the text and stores the translation units in a database while translating. Both the source text and the translation are stored in the database, and are immediately a part of the database - the translation memory – and can be consulted (Dragsted 2003:88). This approach produces high-quality translation memories but it takes a considerable amount of time to build a database of significant size using this approach. Multiple translators can however contribute to a TM, and thus it can be built faster. This method does not allow the translator to exploit a previously made translation, but she must start from scratch (Bowker 2002:109). As opposed to interactive translation, post-translation alignment allows the translator to store previously translated texts into the database and to align them with their translations using an alignment tool which often is supplied with many translation memory software packages (Dragsted 2004:88). The alignment tool then examines the source and target texts and determines which segments belong together. An aligned database is thus built (Austermühl 2001: 135).

Although TM software offer the option of “batch translation”, i.e. the software automatically replaces source text segments with corresponding segments from the translation memory, in reality most texts need to be translated in the interactive mode as HAMT where the translator works assisted by the TM.

The usual set-up when working interactively with a TM system is to have two or three windows open on the computer screen: a text editor, the TM software, and sometimes a terminology software. Most TM software work in conjunction with a word processor. The TM software shows the segment that is being translated as well as possible

matching translated segments. The actual translation is done in the word processor (Dragsted 2004: 89). The translation memory used in this study, Wordfast Pro, is an exception to this rule; it is a standalone platform for which no word processor is needed. More on Wordfast Pro in chapter 4.1.

2.2.2 Terminology tools

There are various ways of staying up to date with terminology. These include file cards, lists in word processors, spreadsheets, database management systems, terminology management systems, and hypermedia systems (Austermühl 2001: 103). File cards are helplessly out of date and fairly useless compared to computerised systems. Creating lists with a word processor or a spreadsheet software also seems arduous and time-consuming, and the search and sort functions are not very versatile. More sophisticated systems are available, such as database management systems. They provide a variety of features and cover different areas, so that beside terminology management the translator can for example store and maintain client addresses and manage orders and projects. Terminology management systems are suited to the particular needs of translators and concentrate on the functions relevant to them, for example importing data and looking up terms. Hypermedia systems are mainly used for data representation, but the input of data can be time-consuming (Austermühl 2001: 105-107).

2.2.3 Corpora

“Corpus linguistics ... investigates relations between frequency and typicality, and instance and norm. It aims at a theory of the typical, on the grounds that this has to be the basis of interpreting what is attested but unusual.” (Michael Stubbs 2001: 151 quoted in Olohan 2004: 16). A corpus is in essence a store of texts, a collection of examples of language, stored in electronic format. The corpus can be then consulted with various search algorithms to gather linguistic information. This separates corpora from traditional libraries and text archives: the aim is not to store texts because of their

intrinsic value, but because of their potential in studying linguistics and linguistic phenomena; thus they are useful for translators as well, for example in examining potential translation equivalents in their contexts (Hunston 2002: 2).

The new paradigm in translation studies, corpus-based translation studies, is according to Laviosa (2003: 45)

> *the branch of the discipline that uses corpora of original and/or translated text for the empirical study of the product and process of translation, the elaboration of theoretical constructs, and the training of translators. CTS makes use of a rigorous and flexible methodology, theoretical principles are firmly based on empirical observations, it uses both inductive and deductive approaches to the investigation of translation and translating, and it encourages dialogue and co-operation between theoretical, empirical, and applied researchers.*

The use of corpora in translation was first suggested by Mona Baker in 1993 (235). Corpora, however, have been used in linguistics for a long time, so it is surprising that it was only some fifteen years ago that the first forays into translation were made. It is, however, understandable if we consider that translation studies did not emerge as a separate discipline until the 1980's, therefore the whole field of study is fairly new as well (Granger 2003:18). Baker (1993:234) and Olohan (2004:13) point out that translations' status within linguistics and corpus studies had traditionally been that of an "outsider": not exactly considered on par with native language but rather an entity of its own, separate from other language use, and therefore perhaps less fertile to investigate.

Corpus access software rearranges the material in the corpus so that various observations can be made. The most common ways of processing data in a corpus are showing frequency, phraseology, and collocation. These terms will be explained in the next three paragraphs.

All the words in a corpus can be arranged in terms of their frequency in the corpus. Usually the most common words in any corpus are grammar words such as *the, of, to, and,* etc. and excluding these leaves the lexical words which reflect the subject matter of the corpus, whether it is for example a collection of political texts or scientific reports. A comparison of frequencies in different corpora can produce various results (Hunston 2002:3).

Phraseology is observed through concordances using a concordancing program. A concordance is a certain word searched in the corpus, and a concordance line shows the word in context, i.e. the surrounding words and sentences are presented as well. A concordance tool shows all the instances of the word in their context, which creates a good overall picture of the usage of the word (Hunston 2002: 9). This is a particularly helpful tool for translators.

Collocation works in a similar manner as concordancing, but collocation produces the actual collocates ("neighbours") of a word: the words that occur in the immediate vicinity of the searched term. Collocation can reveal pairs of lexical items, i.e. words that are used in conjunction with each other, such as shed + tears (Hunston 2002: 12). Concordance and collocation tools are of very useful to translators as they do not treat words as separate, unconnected units but show them in context, which helps translators to deduce the most common usage of a word and its relation to other words and phrases.

Hunston (2002: 14-16) outlines the commonly used corpora types: they are specialised, general, comparable, parallel, historical or diachronic and monitor corpora. A brief overview of each is given next.

- A specialised corpus is a collection of texts of a certain type, e.g. academic articles, political texts, tourism brochures etc. and is used to investigate this particular type of language.
- A general corpus, or sometimes called a reference corpus, is usually a large database of various types of texts, aiming to create an overview of language. Other corpora are often compared against a general corpus.

- Comparable corpora are two or more corpora in different languages or language varieties consisting of similar text types. Translators most often use comparable corpora to check possible translation equivalents and to identify differences.
- Parallel corpora are two or more corpora in different languages, consisting of texts translated from one language into the other or the same texts produced simultaneously in different languages. These are again a good tool for translators and learners to investigate equivalence and differences between languages.
- A historical or a diachronic corpus includes, according to its name, texts from different periods of time.
- A monitor corpus tracks the current changes in a language. It is updated regularly to keep track of the changes.

The terminology can be slightly different in translation studies. Translation researchers often use the term translation corpus to refer to a corpus of translated texts (Granger 2003: 20).

As outlined above, Hunston (2002: 123) claims that the most important corpora for translators are comparable and parallel corpora. They are useful tools for investigating how different terms and phrases have been translated before and in what kind of contexts they occur in language, and how they are used. Corpora can also be used in researching differences between translated language and original language.

There are advantages as well as disadvantages to using corpora. Corpora can reveal a lot of linguistic information that intuition alone cannot reveal, for example judgements about collocation and frequency as well as issues linked to semantics and pragmatics. On the other hand, a corpus is limited by itself. It can only give information about its contents and conclusions drawn from corpora should be treated as deductions only. Most importantly, a corpus represents language out of its original, physical and social context (Hunston 2002: 23).

2.3 Localisation

Localisation refers to the translation and adaptation of a software or web product, which includes the software itself and all related product documentation (Esselink 2000: 1), but many different definitions exist. LISA (The Localization Industry Standards Association) defines localisation as follows:

> *"Localization involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold."*
> (http://www.lisa.org)

Esselink (2000:2) claims that besides the traditional activities related to translation, such as terminology research or proofreading, localisation encompasses a variety of other activities: multilingual project management, software and online help engineering and testing, conversion of translated documentation to other formats, translation memory alignment and management, multilingual product support and translation strategy consulting. Another difference between traditional translation and localisation is that traditionally, translations are done after the source document has been written and finalised. Localisation, on the other hand, often happens simultaneously with product development to allow all language versions to be launched at the same time. This said, it seems that Esselink's view on translation is very narrow; traditional translation projects include a lot of the localization characteristics mentioned above.

Internationalisation is an important factor in localization. According to LISA (http://www.lisa.org)

> *"internationalization is the process of generating a product so that it can handle multiple languages and cultural conventions without the need for re-design.*

Internationalization takes place at the level of program design and document development."

This leads to the two key reasons for internationalising: ensuring that a product is functional and accepted in international markets, and ensuring that a product is localisable (Esselink 2000:25). This also entails that the internationalisation process occurs prior to localisation. At this stage, the translatable text is separated from the software source code to prevent any changes translators may inflict on the source code during the translation process. An essential part of internationalisation is the enabling of local standards in the product, i.e. the ability of the software to display the local character sets and to support foreign keyboard layouts, input methods, and hardware standards (Esselink 2000:3). Globalisation is another important aspect, defined by LISA (http://www.lisa.org) as follows:

"Globalization addresses the business issues associated with taking a product global. In the globalization of high-tech products this involves integrating localization throughout a company, after proper internationalization and product design, as well as marketing, sales and support in the world market."

Globalisation has many uses in different areas, such as economics, business, and internet. Most often it is used in the context of a business expanding its functions to new areas and countries, but in the context of this study it also means making sure that a software or a product is localised to suit the needs of a foreign market (Esselink 2000:4).

Language engineering offers some interesting applications that could be utilized in localisation as well, such as automatic or computer assisted translation, speech recognition and synthesis, speaker verification, semantic searches and information retrieval, text mining and fact extraction. Some of these techniques are applied in CAT tools, which are widely used in the localisation industry. Especially translation memories are useful for working with large volumes of localisation projects and

updating previously translated material. Machine translation tools are also used by some companies. Specifically designed software localisation tools exist which are used to translate and test software user interfaces (Esselink 2000:359,360).

The localisation industry is fairly young. It was only established in the early 80's, and became a prominent field in the 90's along with the growth of internet, which made it easier for businesses to market and distribute their products. Multi-language vendors started to operate and combined the tasks of what is now known as localisation (Esselink 2000:6). The advances made in technology have rapidly changed the localisation industry and continue to do so for the foreseeable future. Universities have also introduced courses in localisation to answer to the demand from the industry (Esselink 2000:6).

2.4 Professional translators and electronic translation tools

Electronic translation tools such as translation memories have become common as translation aids and have only recently been added to universities' curricula, and accordingly, translators trained before the introduction of electronic translation tools into the studies have not received the training and knowledge related to computer-assisted translation.

According to the SPIRIT project (Supporting Peripheral Industries with Realistic Applications of Internet-based Technology) conducted in 2000, a need for usability research exists in the computer-aided translation field. The survey conducted in the study showed that many Finnish translators working in the timber industry rarely used electronic tools other than proofing tools and the internet, and that for example translation memories and corpora were virtually never heard of among the respondents. Many translators found the new technology scary and difficult to learn, and learning to use it seemed to them very time-consuming. Many freelance translators were sceptical about the investment: would the initial investment in an electronic translation tool pay

itself off? The report points out that translators are, however, willing to learn and would like to know more about the electronic tools. The report also illustrates the need to re-develop the tools to be more user-friendly: "…The time-consuming learning phase seems to be a stumbling-block for some users, which appears to point to a need to increase the tool's user-friendliness." (Jääskeläinen and Mauranen, 2000). As Jääskeläinen and Mauranen (2001: 359-360) state, translation tool software seems to be developed by computer specialists or linguists who have not taken the particular needs and wishes of translators into account, and often new tools are published without any knowledge of the above-mentioned wishes.

As García (2006: 97-104) argues, translators see translation memories as both a blessing and a curse: a curse because of the time and energy needed to master the software, and because using a translation memory will encourage clients to demand discounts based on perfect or fuzzy matches. On the other hand it is considered a blessing for the gains in productivity and efficiency when dealing with repetitive texts, and also for the realisation that it can also work well with non-repetitive texts. García notes that translation memories enhance the quality of non-repetitive texts and despite the general preference to use them only with repetitive texts, they should also be considered for other kinds of text types.

A related issue which merits investigating is the continuous training of already working translators. Their IT skills need updating, but there does not seem to exist an effective way of reaching out to all of the translators, or for them to learn the new things in their busy schedule. Courses can be offered or meetings organised, but the key question still remains: how to deliver that information to all interested parties so that they can fully benefit from it, and keep their expertise up-to-date? Some ways of doing this have been proposed. Biau Gil (2006:89) suggests that a 100% online approach could be used, but admits that it does not fully replace the benefits of a face-to-face environment. For a course he taught, a combination of a web page, mailing list and chat sessions were used. He suggests that students with high computer proficiency benefit the most from online courses, whereas students with low computer skills require and seek more personal

tutoring from the teacher and thus would be more at ease with a blended learning environment including both online lessons and face-to-face meetings. Students' initial computer literacy is an important factor.

As knowledge of electronic translation tools is an integral part of a modern translator's competence, it is crucial to include computer technology training in the translation training programmes. Experiences from the COLC project (Computing for Language Careers, 2000-2004) at the Savonlinna School of Translation Studies were encouraging: both students and staff felt that the computer technology training was important and that it should be continued in the future as well. During the project, a new minor subject was introduced: Information technology for translators. It comprised of basic and subject studies of various computer technology courses ranging from localisation to project management. The project improved the public profile of the department locally and in the translation studies field, and both students and teaching staff thought that computer technology should be kept in the curriculum. However, in the end the new minor subject was dropped because participants were not willing to put the time and effort into learning the new technologies (Jaatinen and Jääskeläinen 2006).

3. Usability research

Usability research revolves around the evaluation of user interfaces and their design. The user interface is the only part of the software that the user comes in contact with, therefore it is crucial that it be easy to use and understand. As Ravden and Johnson (1989: 15) point out, a poorly designed interface can lead to misunderstandings, errors, confusion and frustration. It can severely restrict the user's ability to use and understand the system, not to mention lead to serious difficulty in performing tasks using the system. Knowing the end user is crucial, as developing a usable product requires an understanding of the people who represent the actual or potential users of the product. Usability affects the whole product: the hardware, software, menus, icons, messages, manuals, quick references, online help, and training (Dumas and Redish 1999:5-6).

The key notion in usability research is user friendliness, which can be defined in a multitude of ways. As with any software, user friendliness is a major issue with computer-aided translation software. This notion of user-friendliness includes issues such as how easy it is to install the software and to get it running, how seamlessly it can be integrated into the word-processing package already in use and, above all, whether it does what the user expects it to do. Other issues include the quality of the documentation: how easy is it to find out how to do something and/or what the particular function of any part of the software is (Somers 2003: 42).

This chapter introduces the basics of usability research and the basic principles of testing the usability of a software and its user interface. A usability test utilising these principles will be carried out, as described in chapter 4.

3.1 The definition of usability

The definition of usability is sometimes elusive and countless attempts have been made to determine it, but there is no consensus as to which one is the "right" definition (Ovaska et al. 2005: 3). Usability should not be confused with usefulness, which is related, but very different in its relationship to products (Byrne 2006: 97, Dumas and Redish 1999:5). The distinction between different aspects of usability is illustrated in the graph below.

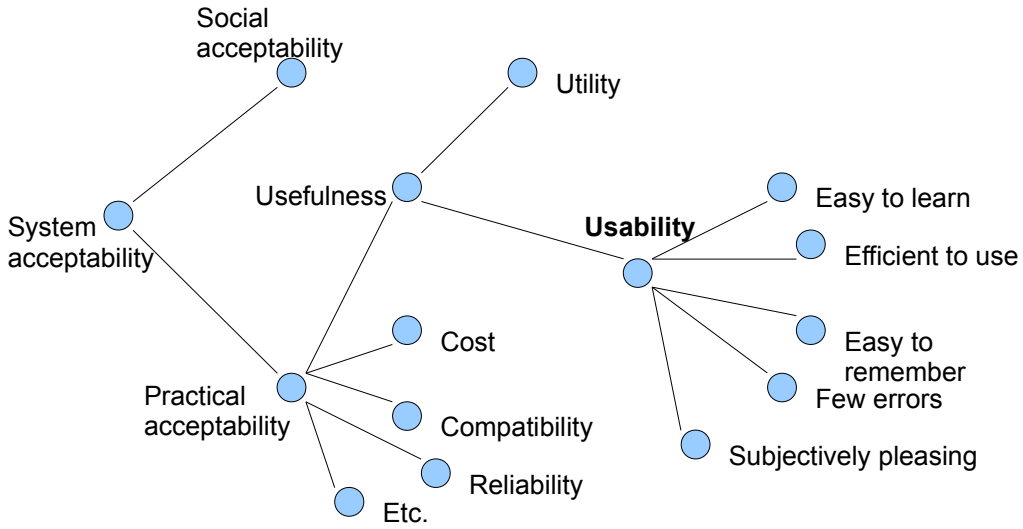


Figure 1. A model of the attributes of system acceptability. (Nielsen 1993:25)

According to Nielsen (1993:24)

> *"usefulness is the issue of whether the system can be used to achieve some desired goal. It can be broken down into the two categories of utility and usability [Grudin 1992], where utility is the question whether the functionality of the system in principle can do what is needed, and usability is the question of how well users can use that functionality."*

According to the ISO 9241-11 standard, usability is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context" (quoted in Byrne 2006: 98). This is a strictly specified view, and Dumas and Redish (1999:4) soften it slightly: "Usability means that the *people who use the product* can do so *quickly and easily* to accomplish *their own tasks*." [emphasis in original]. Ravden and Johnson (1989:9) take a similar view: they define usability as "the extent to which an end-user is able to carry out required tasks successfully, and without difficulty, using the application.

The pioneer in usability research, Jakob Nielsen, lists five main usability attributes (1993:25):

Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?
Efficiency: Once users have learned the design, how quickly can they perform tasks?
Memorability: When users return to the design after a period of not using it, how easily can they re-establish proficiency?
Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
Satisfaction: How pleasant is it to use the design?

These measurable and more precise attributes are needed to complement the elusive term "usability" and to actually be able to form a discipline that systematically approaches, improves and evaluates usability (Nielsen 1993:25).

It should be noted that the experience of usability is in the end the subjective experience of a user, tied to not only the user's personal characteristics but also the task at hand, the equipment and the environment (Ovaska et al. 2005:4). If usability is defined as the user's feeling of success, knowledge of user experience is needed in addition to information about the problems a user encounters while using the system. Surveys and questionnaires are the most common procedure to collect information about user experience (Ovaska et al. 2005:4).

3.2 Usability engineering

Building usability, also called usability engineering, is a long process, spanning from the design and planning of the product to the development and trial stages (Nielsen 1993:71). Usability is not something that can be applied at the last minute, it is something that needs to be built in from the beginning (Dumas and Redish 1999:8). This is accomplished through iterative design which involves end-users throughout the process of designing and building the product. As Dumas and Redish (1998:8) state, usability engineering "starts with identifying users, analyzing tasks, and setting usability

specifications, moves on through developing and testing prototypes, and continues through iterative cycles of development and testing." Quantitative usability goals should be set early in the developing process, for instance "users will be able to understand and set up this software program in less than 10 minutes", thus making sure that designers work to meet these goals (Dumas and Redish 1999:11).

The focus in this study is on methods that collect information about users and usability. These methods are usually divided into two groups: inspection methods and user testing; depending on the user's participation in the process (Ovaska et al. 2005:6). The usability test carried out for this study exploits two different user testing methods: a usability test and a questionnaire.

The key to usability research is knowing the end user and the user's needs. User centred design (UCD) ensures that the defining, designing and evaluation processes during the development stage are repeated enough times to result in products that both the production company and the users are happy with (Ovaska et al. 2005:9). Parallel to this concept, participatory design involves actual users of the product giving their opinions at different stages of the development process (Nielsen 1993: 88). Nielsen (1993:73) also notes that the concept of "user" should be defined to include all those whose work is affected by the product in some way. He emphasises that analysing the background, preferences, age and all other individual user characteristics is essential to developing a usable product.

Task analysis is a tool aimed at discovering the ways in which users perform the tasks that they are involved with now and their future goals, along with their information needs and strategies for different situations and tasks. Identifying users' jobs and tasks helps to design the product with users in mind (Nielsen 1993:75). Task analysis is also essential in planning the tasks that participants will perform during a usability test (Ravden and Johnson 1989: 78).

Competitive analysis should be performed comparing the product with its competitors

on the market. Heuristic analysis of the competing product will give the developers new ideas and information about how the other product works. It is also necessary to map the areas which need more emphasis than others, for example if the focus is on quick learnability. Knowing the strengths and weaknesses of the competitors helps to improve the product (Dumas and Redish 1999:48). Setting the usability goals can be done at this stage, for example determining what amount of user errors is acceptable in performing a particular task. (Nielsen 1993:76-80)

Financial impact analysis entails calculating the estimated loaded cost of the product for the customer as well as the company, i.e. all the costs involved such as the user's salary and other costs like taxes and fees, as well as overhead charges like rent. This will provide an insight into the productivity and profitability of the product (Nielsen 1993:82).

To assess the severity of the usability problems that have been discovered in the unreleased product, companies often ask usability experts to evaluate the problems and form severity ratings which reveal how bad they consider the problems to be. This then can give the developers ideas in what needs to be repaired or edited in the program. Expert-devised heuristic evaluations can also perform the same task if no user test data is available.

It is important that the user interface is understandable to the user. This does not just mean language itself; it is obviously easier to understand a system that uses the user's mother tongue. It also entails that the language is appropriate for the target users, i.e. if the users are not specialists in a certain area the software should not use specialist terminology that the users may not understand, and vice versa: subject-field terminology can be included in specialist software (Nielsen 1993: 123).

User manuals and other documentation are also a consideration for usability testing. When users do read documentation (most prefer to spend their time on more “proactive” activities, i.e. performing tasks on the system), they expect to find what they are looking

for quickly and easily. Whether the help documentation is online or in a printed format, it should easily accessible, easy to use and easy to understand as well as cover all the topics needed (Nielsen 1993: 149).

Dumas and Redish (1999: 64) state that to assess the severity of the usability problems that have been discovered in the unreleased product, companies often ask usability experts to evaluate the problems and to form severity ratings which reveal how bad they consider the problems to be. This can give the developers ideas in what needs to be repaired or edited in the software. Expert-devised heuristic evaluations can also perform the same task if no user test data is available. According to Nielsen (1993:155) heuristic evaluation is "done by looking at an interface and trying to come up with an opinion about what is good and bad about the interface" and "...a systematic inspection of a user interface design for usability.... Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognised usability principles (the "heuristics")." This results in a set of problems uncovered by the evaluators, which then can be remedied. Heuristic evaluation is not guaranteed to reveal every single usability deficit in the product, but it is an inexpensive method that will uncover all the major usability issues, depending on the experience of the evaluators (Nielsen 1993: 155-163). Heuristic evaluation is a more analytical approach compared to usability testing, which is more of a hands-on test method (Sharp et al 2007:592).

Gathering information in the actual environment in which the product is going to be used is usually a part of the process of determining the requirements of the product, i.e. requirements analysis. Laboratory tests usually focus on discovering usability problems (Ovaska et al. 2005:6). In recent years usability research has focused more on user participation, e.g. testing the product in a situation where the user uses the product in an environment that the end product would be used in, and the aim is to understand usability in this very environment in various situations and contexts. User participation has even been integrated into the development process (Ovaska et al. 2005:7).

As this is a student paper and not related to the production company behind WordFast,

the test was performed after the product had already been developed, finalised and published. In an ideal situation, these tests are carried out before publication to iron out all the bumps. The participants of the usability test are all students, and not professional translators. They are, however, representative of the target user group to a great extent because CAT training is included in their studies and all have at least some experience using them, if not on a professional level.

Evaluating user interfaces is important not only in terms of usability and functionality, but also user satisfaction. Users not only look for a usable system, they also want a pleasing and engaging experience. Other reasons for testing include designers getting feedback on their early design ideas, problems being corrected before publication and also, with testing, the focus can be put on the real problems instead of superficial issues (Sharp et al. 2007:586). Not doing any testing can also lead to for example omissions of the "right" features and inclusion of "wrong" features; arguing on the basis of no data; time and money spent redesigning inconsistent or unusable user interfaces; an increased number of support calls, and negative impressions of the product and loss of sales (Wixon and Wilson 1997:659).

Various methods of usability research exist but it is beyond the scope of this study to go through all of them, therefore only the methods relevant to this study are presented in the following chapter. These methods are usability testing and questionnaires.

3.3 Usability research in the translation industry

Although a field not touched on very often in translation studies, usability research is nevertheless as important to translators as it is to other computer users. Monika Höge (2002) studied the evaluation of translator's aids from an interdisciplinary point of view combining translation theory, software engineering and decision analysis to evaluate which translation tool was the most suitable for a specified task. She came to the conclusion that combining these disciplines provided an extensive understanding of the

processes included in evaluating electronic translation aids.

Höge examined the problems that occurred during the process of translation and what techniques translators used when working with texts. Although this provided some insight into features that would be useful when transferring a text from one language to another, it could not provide a detailed analysis of them. Therefore she applied requirements analysis to elicit the parameters relevant in the translation context. Next, she applied the ISO 9126 quality tree to evaluate translators' aids to help develop the system properties and to establish an assessment function for the particular context. She found out that when considering which tool would be most useful, the results from a number of testing procedures were applicable and scenario tests could show whether the figures obtained during various stages of testing really make sense in the practical working environment.

As Höge points out (2002:1), “research in this area is impeded by its interdisciplinary nature and the requirement for integrating findings from different scientific areas such as software engineering, decision analysis, and translation.” She notes that software engineering has a significant impact on the development of software evaluation methodology. Decision analysis procedures can be applied for evaluation purposes and share a great detail of characteristics with it. The most important factor, translation, serves as a basis for understanding the specific issues related to evaluating translators' aids. Her evaluation method works as a cycle: it first examines the user and the system, followed by the modelling stage which determines the relevant attributes and test types. Next the attributes are tested, and finally analysed. Test results are validated and brought back to the user.

The issue with Höge's study is that it is already somewhat dated, as the starting point for it was the CAT tool market of 1999/2000, and the author describes the reality as translators receiving their work as print outs or on floppy disks. Now, more than ten years later, much progress has happened both in technology and in the translation tool market and the approach in the study may not be fully applicable any longer.

The obvious goal of evaluating translators' aids is to help support translation work (Höge 2002:3). A deep understanding of the translation process and relevant problems is required to properly evaluate the usefulness and usability of translation tools. Höge claims that the main issue is that up until now, translation has been considered an isolated activity, whereas she sees it as a part of documentation life cycle. Moreover, she argues that when choosing the right translation tools, the focus should be not only on the problems that the translator faces but also on the technical background, the original author's strategies, the final users of the document, and the context in which the translation is carried out. According to her these factors will help determine where, when and what translators' aids' systems actually facilitate translation.

Although Höge's study does not strictly concern usability testing, she raises many issues that usability testing shares. The need to understand the particular needs of the end users, in this case translators, as well as the process of translation, is a universal requirement. Usability testing in the translation industry should ideally combine elements from different fields of study, much like Höge's study, to provide the testers with a full understanding of the factors affecting the process of translation and also the process of usability testing. Höge (2002:7) argues that the need for a framework for evaluating translators' aids arises from two issues. These are:

(i) In the translation industry, there's a need to assess the system before purchasing. Translators face the pressure from shortening product life-cycles and the increase in international communication, and therefore do not feel they have the time to implement new software into their systems, and do not have the methods to asses the alternatives in terms of costs and benefits.

(ii) In translation system development, supporting the development process. The development of translators' aids' systems requires computer-based knowledge of the particular issues faced during translation.

These two points can be applied to usability testing in the translation industry as well.

Efforts have been made in the past at forming guidelines for evaluating translation tools, with emphasis on different areas of importance and with varying outcomes, such as Thompson 1991 and 1992, or Kugler/Ahmad/Thurmair 1995. An important study in forming a general framework for evaluating natural language processing tools in general was made by the EAGLES evaluation and assessment group. According to Höge (2002: 38)

> *"EAGLES was intended to respond to the lack of common technologies and standards for the language industries. From a practical point of view, the major objectives of EAGLES were*
> *(i) to produce agreed specifications and guidelines for specific areas of language engineering and make recommendations for a more uniform approach; and*
> *(ii) to bring together the different approaches of industry and academia and foster their collaboration. Five main areas were identified to form working groups:*
>
> *(1) Text Corpora*
> *(2) Computational Lexicons*
> *(3) Linguistic Formalisms*
> *(4) Evaluation and Assessment*
> *(5) Spoken Language Resources and Methods*
>
> *The working group on evaluation and assessment was split up into three subgroups, that is, Writer's Aids, Translators' aids and Information Management Systems. Jointly the three groups strove to set up guidelines for the evaluation of language engineering products and to exemplify the validity of the guidelines by applying them in the three areas of interest."*

EAGLES developed a division into three types of evaluation: adequacy evaluation which evaluates a system in respect of some intended use of that system, progress evaluation which evaluates a system in respect some desired state of the system, and

diagnostic evaluation, which evaluates the system to find out where it fails and why (Höge 2002:38). The most adequate form of evaluation for the purpose of the present study would be adequacy evaluation: the intention is to find out how well a translation memory system works when it is used for its intended purpose – translating. The aim is to find out if the system will do what is required and how well it does it and at what cost; most often this evaluation method would precede purchase decisions. However, the final report of EAGLES could not determine definite methods for evaluating translators' aids' systems in terms of user needs.

Leena Salmi (2002:84) suggests using log files, the think-aloud method and pair work in usability tests. Log files recorded by the computer can produce a lot of quantitative data about the detailed use of the system. The think-aloud method works for solitary users, who are asked to vocalise their thought processes during testing. This method helps the researchers to understand how the users view the computer system. However, the thinking-aloud situation may not feel natural, and to counteract that, Salmi proposes pair work as a viable alternative. Using this method, the participants discuss and solve problems in pairs, producing information about human-computer interaction.

To determine what the translator wants from the system and how the system's functionality corresponds with these wants is a vital part of the evaluation process, and an element borrowed from requirements engineering. To determine how useful a particular system is for a particular translator, the specifications of both have to be mapped onto each other to discover the attributes that are relevant to both parties. These shared attributes can then be evaluated in terms of user needs and how well the machine corresponds with them (Höge 2002:44-46). The same logic can be applied to usability testing. More research in the particular field of translation is needed.

3.4 Questionnaires

Questionnaires or surveys, as well as interviews, are mostly used to collect data about

the respondents' thoughts, opinions, and feelings, which are difficult to measure objectively (Nielsen 1993:209). Questionnaires enable the researchers to gather large or small amounts of data, and can be modified to suit various kinds of research projects. They can be used in all stages of the development process of a product. They can form the sole method of information gathering, or be a supplementary method in a project. The particular details of a project determine the suitability of questionnaires as a data collection method for that particular study (Vanhala 2005:11,17). All questionnaires should be subjected to pilot tests before actual use in testing to avoid ambiguity and anything that could lead to misunderstandings. Effort and skill are needed to form unambiguous questionnaires that produce data that can be analysed efficiently (Sharp et al. 2007:308). Questionnaires can include open questions, but closed questions (i.e. when users have to supply a single fact, tick off answers in a list, or rate something on a scale) are easier to interpret and analyse and therefore are more common (Nielsen 1993:212).

Usually questionnaires are applied to gather a large amount of data from a variety of respondents. In this study, as the number of respondents is fairly low, the same participants who take part in the usability test will be presented with a questionnaire after the test to collect their thoughts and feelings on the tasks and on the system.

Commercial questionnaires for usability testing that everyone can utilise (for a fee, most often) are available, mostly in English. The QUIS (Questionnaire for user interaction satisfaction) questionnaire is an example of a ready-made questionnaire frequently used for evaluating interfaces, and as it has gone through several cycles of testing and refinement it is well tried and tested. The QUIS measures system experience, past experience, overall user reactions, screen design, terminology and system information, learning, system capabilities, technical manuals and online help, online tutorials, multimedia, teleconferencing, and software installation; however many testers only choose to use the section that measures users' overall reactions (Sharp et al. 2007:310). Ready-made questionnaires may not, however, offer enough detailed information about the subject of study, the product in question, so it is advisable to customise them or

design a questionnaire of your own. Creating a questionnaire from scratch is a very delicate process that takes both time and dedication, and requires repeated testing and modification (Vanhala 2005:17-18,22).

The reasons for modifying an existing questionnaire can be manifold. As stated above, it may not give enough detailed information about the issues the researcher may wish to uncover. The ready-made questionnaires can also include a lot of unnecessary questions that can frustrate the respondent. The length of the questionnaire, the differences in the intended target group and target audience of the questionnaire, language and design of the questionnaire can all affect the outcome. The data collection method can also be different than that of the original (Kitchenham and Pfleeger 2002c:20)

As Vanhala (2005:25-33) notes, designing a questionnaire from scratch is an industrious task, but may be necessary if ready-made questionnaires do not directly answer questions that need answers or do not seem to directly measure the qualities of the product or issues at hand. Questions can have two answering methods: open questions or ready-made answer options (multiple choice). Multiple choice produces results that are easier to analyse and collect; multiple choice produces quantitative material whereas open questions produce qualitative material. A common answering pattern for multiple choice questions is the Likert scale, or other scales of the same kind. The scales consist of five or some other amount of numbers where one extreme is “Strongly disagree” and the other “Strongly agree”, or something similar such as “Never” and “Always”. Another possible scale variant is the semantic differential scale, which presents the user with bipolar word pairs and asks the user the place a cross on a number of positions between the two opposites, such as clear.....confusing. These are however less popular as it may be difficult to think of enough word pairs (Sharp et al. 2007: 315). The reliability of the answers and the results of the questionnaire can be judged by conducting the same questionnaire many times with the same respondents and the same questions. Careful design of the questionnaire, including many phases of testing, produces results of high accuracy. Choosing the right sample (e.g. respondents) for the study can be crucial depending on the subject of research, as well as understanding the

reasons behind answering patterns. The time at which the questionnaire is completed also plays a role: according to Hufnagel and Conca (1994, quoted in Vanhala 2005:33) if respondents are asked to answer questions during the test or between different tasks, they are more prone to describe the issues and problems in more detail than after completing the whole test and all tasks (Vanhala 2005:25-33).

The data gathered from questionnaires is always subjective, and it may not be easy to identify the factors affecting the judgements. Moreover, it could be easier to measure the subjective feelings of the respondent than the objective qualities of the product (Vanhala 2005:20).

3.5 Logging

Logging involves the computer automatically collecting detailed data and statistics about the use of the system. Typically, it collects frequency data about how often tasks were performed by users and how often they encountered errors or other problems. Logging can reveal various things, such as which commands and functions the users do and do not use, along with error data. Logging is most often done by instrumenting low-level part of the computer system such as keyboard and mouse (Nielsen 1993:217-18). In a usability test interaction logging, as Sharp et al. (2007:340) call it, is most often synchronised with video and audio logs to facilitate the evaluators to understand users' behaviour and actions.

3.6 Usability testing

Usability testing is a term used to refer to tasks performed by target group users using the product in situations resembling the authentic environment for product use. It is a systematic way of observing actual users using the product and collecting information about the specific ways in which the product is easy or difficult for them to use (Dumas and Redish 1999:12). The goal is to discover which qualities of the product do not work

as they are supposed to or need remodelling. The focus is on the target user. Usability testing often requires more resources than other usability research methods, and it is time-consuming (Koskinen 2005:187). It is, however, the most fundamental usability method.

According to Dumas and Redish (1999:22), every usability test, no matter where and how it is conducted, shares these five characteristics:

1. The primary goal is to improve the usability of the product. For each test, you also have more specific goals and concerns that you articulate when planning the test.
2. The participants represent real users.
3. The participants do real tasks.
4. You observe and record what participants do and say.
5. You analyse the data, diagnose real problems, and recommend changes to fix these problems.

Usability testing is only successful when it leads to improvement of a product and corrections in some parts of it. The term "usability testing" is often used to refer to various techniques or methods of usability research, but here it is used strictly to refer to the actual testing process involving users, such as defined by Rubin (1994: 25): "a process that employs participants who are representative of the target population to evaluate the degree to which a product meets specific usability criteria". It is also important to notice that the terms "usability" and "testing" can also be vague and misleading, especially to someone who is not an expert on the subject, and even in the usability research field consensus has not been reached as to the consistent usage of these terms (Koskinen 2005:188).

In usability testing situations information about the product is gathered through recording and observing the behaviour and actions of target users while they use the product in question. There can be anything from one to hundreds of testers depending on the product and the test. The normal set-up for a usability test is one user and from

one to three observers, some of whom can be hidden from the user. The user is asked to perform a series of tasks resembling real life tasks performed with the product. Information collected is usually in the forms of notes, video and written logs, and these are then analysed. The test can focus on the whole product or on a part of it. Usual subjects for usability testing include for example computer software, internet pages, mobile services and electrical products. The testing is most often carried out by the usability research department of the mother company, or at an outside institution specialising in usability testing such as universities or businesses (Koskinen 2005:188).

The obvious goal for usability testing is to discover and repair usability deficiencies in a product and its supporting materials, but motives for testing can span from a variety of sources. Rubin (1994: 26) names a few, such as companies wanting to create a record of their past usability tests to be able to keep up to the standards in the future, minimising the costs of customer support through a product that works and is easy to use, increasing present and future sales as happy customers return to the same product and recommend it to others, acquiring a positive image as an “easy-to-use product”, and avoiding the risk of publishing a product that has serious usability deficiencies. Dumas and Redish (1999:17) claim that the savings yielded by these measures should be greater than the expenses.

Usability tests normally provide the following types of data, or some of it depending on the methods used, according to Wixon and Wilson (1997:666): time to complete a task; time to complete a task after a specified time away from the product; number and type of errors per task; number of errors per unit of time; number of navigations to online help or manuals; number of users making a particular error; and number of users completing a task successfully. Software logging records for instance keystrokes and mouse movements and together with video recordings form a method for inspecting the user's actions very closely.

Ravden and Johnson (1989: 19) suggest using a nine-point check list to evaluate an interface after performing tasks with the system. These points are:

1. visual clarity (information displayed on the screen should be clear, well-organised, unambiguous and easy to read);
2. consistency (the way the system looks and works should be consistent at all times);
3. compatibility (the way the system looks and works should be compatible with user conventions and expectations);
4. informative feedback (users should be given clear, informative feedback on where they are in the system, what actions they have taken, whether these actions have been successful and what actions should be taken next);
5. explicitness (the way the system works and is structured should be clear to the user);
6. appropriate functionality (the system should meet the needs and requirements of users when carrying out tasks);
7. flexibility and control (the interface should be sufficiently flexible in structure, in the way information is presented and in terms of what the user can do, to suit the needs and requirements of all users, and to allow them to feel in control of the system);
8. error prevention and correction (the system should be designed to minimize the possibility of user error, with inbuilt facilities for detecting and handling those which do occur; users should be able to check their inputs and correct errors, or potential error situations before the input is processed); and
9. user guidance and support (informative, easy-to-use and relevant guidance and support should be provided, both on the computer (via online help facility) and in hard-copy document form, to help the user understand and use the system.

On top of these nine points Ravden and Johnson add two sections about the general usability issues:

10. system usability problems; and
11. general questions on system usability, which address the overall usability issues in different parts of the system.

Based on these 10 sections, the evaluator should be able to form an overall picture of the usability of the system in question as well as detailed information about particular problem areas. This method also reveals the good aspects of the software and how well the system meets each criteria.

Testing is not, however, without limitations. As Rubin (1994:27) explains, it does not guarantee that a product will be usable or that every usability problem is identified. Moreover, testing is always performed in an artificial situation, whether in a laboratory or in the field. The test results are not proof of a working product as they are very much related to the way in which the test was conducted. Participants in the test often do not fully represent the target group because defining a target user is difficult and the participants only represent the target users to the extent of the test conductors' view of them. The pitfalls of testing are to do with the reliability and validity: individual differences between users and the actual merit of the results as indicative of real usability issues (Nielsen 1993:165-169).

The research design of this study is based on the points raised above. The plan is fairly straightforward: to recruit participants, ask them to test the software and fill a questionnaire. The particulars of usability testing are described in the next chapter.

3.7 Conducting usability research

3.7.1 Planning the usability test

It is necessary to carefully consider the aims and the methods of the usability test beforehand and develop a test plan. Clear objectives for testing should be drawn, taking into consideration the available resources. At least the following should be taken into account when planning the test:

- The goals of testing. There must be a reason behind organising a usability test, for example an effort to increase the usability of the product.
- Usability problems to be solved and aims to be reached. After finalising the goals of testing, the focus shifts to the questions that are going to be emphasised during the test. At this stage clear and precise questions about the product should be formed, such as “Does the user understand the function of each button without consulting the manual?”. For example “Is the product usable?” is too vague a question.

- Target groups. If the participants of the usability test do not represent the actual target group of the product, testing is most likely fruitless. It is crucial to know what the end users of the product are going to be like and then recruit the right participants.
- Every detail of the test arrangements should be considered in advance. The proceedings and methods of testing should be planned. It is not until the test plan reveals what it is that the test aims to uncover of the product and its users that the methods and the materials can be decided upon.
- Drawing the test tasks. This is the most critical stage of the test and it is crucial to understand the context of usage and the needs of real users.
- Preparing the test environment and equipment. The best environment for a usability test resembles the real environment in which the product would be used.
- Monitoring of the test should be arranged. Monitors can have different roles in the test.
- Collecting and analysing the results/evaluation measures.
- The content of the final report.
- Test budget.

(Koskinen 2005:189-192, Rubin 1994:83, Nielsen 1993:171, Dumas and Redish 1999:106)

A test plan is the basis of the whole test. It covers every detail of the test and should be followed throughout. It serves as a channel of communication between the developers, monitors and everyone else involved in the process. It states the required resources and provides a focal point for the research (Rubin 1994:81-83).

3.7.2 Selecting participants and user characterisation

The number of participants can be anything from a few to dozens, depending for example on the validity of results required, the available resources, the available

participants and the time required for the preparation and the test (Rubin 1994:128). The usual guideline is “You cannot have too many participants”, and for statistically valid results a minimum of ten to twelve participants per condition should be used. It is however perfectly possible to conduct a less formal usability test with just four to five participants. This number of participants is shown to expose 80 percent of the major usability deficiencies in a product (Virzi 1990, quoted in Rubin 1994:93 and in Dumas and Redish 1999:127). It is important that the participants represent the intended end user of the product as closely as possible, otherwise the results of the test may be dubious or of limited value. The key to selecting the right participants is identifying the relevant skills and knowledge of the end users, i.e. the user profile or user characterisation of the target population (Rubin 1994:120). In the particular case of this study, defining the primary audience of the product is fairly straightforward: translators and other professionals in the translation industry. In light of the SPIRIT project final report (Jääskeläinen and Mauranen 2000:11), which revealed that some translators find it hard to learn how to use electronic translation tools, it is assumed that some translators will find high technology difficult to understand and use. It is however difficult to make any generalisations based on assumptions. In an ideal case, user profiles would be compiled from a variety of information sources such as market research, competitive analysis, focus group sessions, contextual interviews, and observing and interviewing the prospective users (Dumas and Redish 1999:120).

Defining a user profile for a professional in the translation industry is not a straightforward task. Translators come from a variety of backgrounds, and not all have formal training or a degree. Concentrating on the newly graduated or newly started translators, it is easier to generalise: most have a university-level degree and experience in using computers. Newly graduated translators are also more likely to use translation memories and other translation tools in their work than translators from previous generations. As pointed out, the prospective user group is varied, and there are some characteristics that may divide the users, such as experience in using translation memory systems, which can be anything from practically nothing to many years or possibly even decades. This was an important factor in the current study. In this study,

participants with little translation memory experience were used to discover how easy it is for them to learn how to use Wordfast Pro.

Rubin (1994:126-129) states that it may also be a good idea to introduce groups consisting of various skill levels and expertise into the test. This way, all aspects of the product will be used and the different user groups among the users represented. Moreover, Rubin also states that including a "least competent user", LCU, in a test will help ascertain the usability of the product, for if a user with very little knowledge of the product or computers can perform the tasks, then the majority of the end users most likely will as well.

The ideal test subjects for the test in this study would be people of all ages, as young and older people tend to have differing attitudes when it comes to for example computers and technology (Nielsen 1993:176), although this is rapidly changing in modern society. The scope of this study did not, however, allow to search for a larger group of participants and the author had to settle for users readily available at the time of writing: students. As mentioned before, they are in fact to a great extent representative of the target user group apart from the age factor as they all tend to be in the same age bracket, in their twenties.

3.7.3 Other materials needed for the test

As listed by Rubin (1994:141-212), materials that need to be planned and written prior to the test are a screening questionnaire, an orientation script, a background questionnaire, data collection instruments, a non-disclosure agreement and tape consent form, a pretest questionnaire, task scenarios, prerequisite training materials, a post-test questionnaire, and a debriefing topics guide. Some of these have been mentioned earlier but nonetheless a brief overview of each will be given. *A screening questionnaire* is the means for qualifying and selecting the right participants and usually contains parts from the user profile and test plan. Potential participants fill it in before taking part or being

rejected from the test. In this study, a screening questionnaire was not used because participants were recruited through channels that were available to suitable candidates only. *An orientation script* or a test script describes in detail what will happen during the test session and it is read to participants prior to the test. It informs the participants of their tasks. An informal orientation script was written for the current study. *A background questionnaire* is filled by the participants before the test and reveals their experience, attitudes, and preferences in areas related to the test and the product in question. This questionnaire is often a valuable asset in analysing the test results and the answers. The background questionnaire was of great importance in this study. *Data collection instruments* (or data loggers) refer to the ways and tools with which data is collected and analysed from the test. Performance data is objective measures such as error rates and time measures, and preference data is more subjective data collected from the participants' feelings or opinions. There are numerous ways to collect data. In this study, a combination of a background questionnaire, a post-study questionnaire and screen recordings was used. *A non-disclosure agreement* is signed by participants so that no proprietary product information possibly gained during the test will be published. In this study this was unnecessary as the product is already on the market. *A tape consent form* is signed by the participants, giving their assent that the test session can be filmed. A consent for was used in this study for taking screen recordings. *A pretest questionnaire* explores the participants' first impressions of a product and their level of expertise, and can be used to place the participant in a specific group. This study used the background questionnaire for this task, and therefore a separate pretest questionnaire was not used. *Task scenarios* are, according to their name, representations of what the participants will be asked to do during the test, including the contexts for tasks. These were drafted informally for this study. *Prerequisite training materials* are optional and refer to any training given to the participants before the test. In the current study, there was no training. *A post-test questionnaire* gathers information about the participants' opinions of the product after completing the test tasks. This was an important feature of this study. *A debriefing topics guide* lists the subjects which the monitor would like to discuss during the post-test debriefing session. There was no recorded discussion in this study.

Pilot tests should be arranged to try out the test procedures. A pilot study is a small trial run of the main study, aimed to make sure that the proposed method is viable (Sharp et al. 2007:293). A few pilot subjects is normally enough, if it is not a large-scale test. In the pilot test, any ambiguous or hard-to-understand instructions and other faults or mistakes in the procedure should arise (Nielsen 1993:174). If lacking in test users, asking colleagues and other peers can be a quick and easy substitute for a pilot test (Sharp et al. 2007: 294). The usability test planned for this study was subjected to pilot testing: two users completed a pilot. Based on the pilot tests, the post-test questionnaire was radically changed (more on this in chapter 4).

3.7.4 The test situation

A usability test usually has four stages: preparation, introduction, the test itself, and debriefing (Nielsen 1993:187). Preparation involves preparing the test room and the materials as well as the testers themselves (also called monitors or experimenters). The test begins by introducing the test and the testers to the test users and distributing all the necessary handouts, forms and questionnaires. The users normally fill in a form with their personal information, i.e. the background questionnaire, and sometimes a consent form for recording (e.g. videotaping the test situation) is needed. The background questionnaire is used to collect information that can have an important role in the test, such as age, sex, and experience. According to Barnum (2002, quoted in Koskinen 2005: 193) the introduction to the test typically includes:

- introduction to the test room and equipment
- introduction of the monitors
- introduction of the product to be tested
- demonstration of the tasks to be performed during the test
- presentation and demonstration of the think-aloud protocol or other method used
- reminder that the goal is to investigate the product, not the user.

It is also important that the test user knows that they can stop the test at any time and that they are aware they will not receive any guidance on how to use the product during the test (Koskinen 2005:193). The test users should also be given the chance to ask any clarifying questions before the test begins. The monitors should also consider the ethical aspects of testing, such as not making the test subject feel nervous or embarrassed, and keeping the results confidential and anonymous (Nielsen 1993: 181). Nielsen (1993:184) outlines the main ethical considerations:

Before the test:

Have everything ready before the user shows up.

Emphasize that it is the system that is being tested, not the user.

Acknowledge that the software is new and untested, and may have problems.

Let users know that they can stop at any time.

Explain any recording, keystroke logging, or other monitoring that is used.

Tell the user that the results will be kept completely confidential.

Make sure that you have answered all the user's questions before proceeding.

During the test:

Try to give the user an early success experience.

Hand out the test tasks one at a time.

Keep a relaxed atmosphere in the test room, serve coffee and/or have breaks.

Avoid disruptions: Close the door and post a sign on it. Disable telephone.

Never indicate in any way that the user is making mistakes or is too slow.

Minimize the number of observers at the test.

Do not allow the user's management to observe the test.

If necessary, have the experimenter stop the test if it becomes too unpleasant.

After the test:

End by stating that the user has helped you find areas of improvement.
Never report results in such a way that individual users can be identified.
Only show videotapes outside the usability group with the user's permission.

During the debriefing session after the test, the user is asked to fill in the possible questionnaires, and only after that can the discussion begin. Users are typically asked for comments on the software and their suggestions for improving it, as well as any comments on the test itself. After the test user has left, the monitor should write a brief report about the experiment as soon as possible as recent events are easier to remember than older (Nielsen 1993:191).

Some usability testing is conducted in specially-built usability laboratories, but it is not necessary to have one as a test can be carried out with very little equipment, even just a notepad in some cases. A laboratory environment does however facilitate a more private and efficient testing session and provides the test monitors with easier access to the situation via one-way mirrors, video cameras etc. (Nielsen 1993:202).

In this study the users were asked to sign a consent form for recording their actions on screen. For this, a screen recording software called HyperCam was used. Everything that occurred on the screen was recorded, but no sound or keystrokes were recorded.

3.7.5 Evaluation

According to White (2003: 230) the evaluation of machine translation systems may include the following: the time to complete a particular task, the number of steps to complete it, how natural the navigation process appears to be, how easy it is to learn how to use the application and how helpful the documentation is. Dumas and Redish (1999: 185) expand on these basic guidelines. They list the following as the measurable objects in a study in addition to the ones mentioned above: amount of errors, time spent on each screen, the number of steps taken to complete a task, the number of actions

taken to complete a task, the number of incorrect menu choices/icon choices/repeated errors/other errors, time spent in online help or reading a manual, time spent finding information in manual and time spent navigating menus. These basic guidelines can be applied to the present study as well.

In evaluation there is always room for human factor bias and it is important to make sure that the errors are attributed to the usability of the system and not to personal characteristics. This is achieved by getting a large sample of users and controlling the sequence that is measured, and ensuring the testing circumstances are as analogous as possible (White 2003:230). The problem with the current study is the sample size: there were only six participants, which poses questions about validity. As was pointed out in chapter 3.7.2, a minimum of ten to twelve participants is required for statistically valid results. Some difficulties in recruiting participants were experienced, and thus the number of participants is only six. It is not feasible to make generalisations based on such a small sample of tests, but as mentioned in chapter 3.7.2, it is enough to uncover the major usability flaws.

Usability issues will be categorised according to the following scale developed by Nielsen (quoted by Koskinen 2005:198) according to their seriousness:
0 - not really a usability issue, just differences in opinion by the examiners
1 - cosmetic usability issue, will be fixed when and if there's time and money.
2 - minor usability issue. Low priority, will be fixed after more serious issues have been fixed.
3 - Serious usability issue, high priority, should be fixed as soon as possible.
4 - Catastrophic usability issue, prevents the intended use of the application, must be fixed immediately.

Forming the suggestions for improvement after identifying the usability issues entails identifying the parts of the application that cause problems, and then forming the suggestions. Expertise and experience is essential in this task (Koskinen 2005:199).

4. Research material and method

In the usability test in this study, six translation students were asked to translate a short text using Wordfast Pro demo and then answer a questionnaire. No ready translation memory was provided; the test users were asked to start from scratch and create a new translation memory. No help was provided in using the software other than the built-in help section and the internet; the aim was to discover how easy the program would be to take in. The test users were then asked to fill in a post-test questionnaire about their performance and what they regarded difficult or problematic in the translation process. Users were also asked to sign a consent form for screen recording, and to fill in a background questionnaire. A screen recording was made of each test user's performance using the screen capture software HyperCam to better analyse and compare the issues that emerged during the usability test.

The material, i.e. the source text for the usability test, consists of a two-page excerpt from a heat pump/air conditioner user manual, which was obtained from the http://www.sanyo.com website. The excerpt contains instructions on how to use the remote control.

Planning for the usability test started with finding a suitable text and translating an already-existing English usability questionnaire, consisting of categories of multiple choice questions, and drawing the necessary plans as stated in chapter 3.7.3. A pilot test was arranged which two participants completed. After the pilot tests it became clear that the ready-made questionnaire formula did not work for the purposes of this study as it was very long and contained very detailed questions about different parts of the software. A new version of the questionnaire was drawn, this time with open questions. This version of the questionnaire, however, was not subjected to pilot testing due to time restraints, contrary to the general guidelines.

The actual test situations were arranged to take place at either the present author's home or at the university library. First, the test and the methods were introduced. Test

participants were asked to sign a consent form to allow their performance to be recorded and to fill in a background questionnaire stating their age, educational background, computer and translation memory experience, as well as experience in translation work. Then they were asked to translate the text according to the given instructions which were to create a new project and a translation memory and then translate the text, with the fuzziness threshold set at 65. The instructions told them clearly to create a new project and a new translation memory, and then translate the text. It seems some participants failed to read these instructions carefully enough, as will become evident in chapter 5.2.4. After they had finished translating, they were asked to fill in a questionnaire which consisted of open questions about usability and about issues they faced when translating. As was stressed in the beginning of the test, they did not receive any help from the test monitor apart from if technical issues with the computer itself arose. Each experiment was recorded using a screen capture program, HyperCam.

The tests were carried out one at a time using a laptop with Windows Vista Home Premium and the following software:
- Wordfast Professional 2.4.1 demo version
- HyperCam 2.25.01
- Mozilla Firefox 7.0.1.

A total of six participants volunteered for the test. They all finished the assignments more or less successfully, and faced a number of issues depending mostly on their experience with similar translation memory tools and probably, interestingly, on their ability to read instructions. All of the participants were students of English and translation at the University of Eastern Finland and had taken courses in translation, and were therefore acquainted with translation memories. Nearly all of them felt unsure about their skills, however, and felt that they still had much to learn in using translation memories. Two of the participants had some experience in translation work but this did not directly correspond to faster completion times or more effortless use of the software, as will become evident in chapter 5.

The answers in the questionnaires varied greatly; one participant described the tasks as very easy and another as difficult. This is mostly explained by the different backgrounds of the participants. To retain each participant's privacy their name was not asked in conjunction with the test and the test materials, and the materials were only stored on the present author's computer for privacy protection and safety. All of the participants will be referred to as a "she" to preserve anonymity, regardless of their actual gender.

4.1. Wordfast Professional

Wordfast Professional, or Wordfast Pro, is a translation memory tool that can be customised to suit the user's preferences. It is a standalone tool that is not integrated with a word processor, unlike Wordfast Classic. The version used in this study was 2.4.1. The user can alternate between two perspectives of the interface, depending on what they are doing: the PM (project manager) perspective and the TXML Editor perspective. Translations can be carried out in the TXML Editor perspective in two modes: the table mode which shows source segments and target segments side-by-side, or in the text view mode which shows the source and target segments in succession. The PM perspective offers functions such as batch file analysis, clean-up and TM updating. The user can switch between perspectives using the buttons on the left-hand side bar of the interface.

On the TXML Editor, on the top of the interface, there is a long row of icons, some of which are familiar from common use, such as the save and open file commands. The same commands can be found in the drop-down menus, but are situated on the interface for quick access. The command for each icon appears when hovering the cursor over the icon for a few seconds. The following screen capture illustrates the layout of the TXML Editor perspective in table mode.

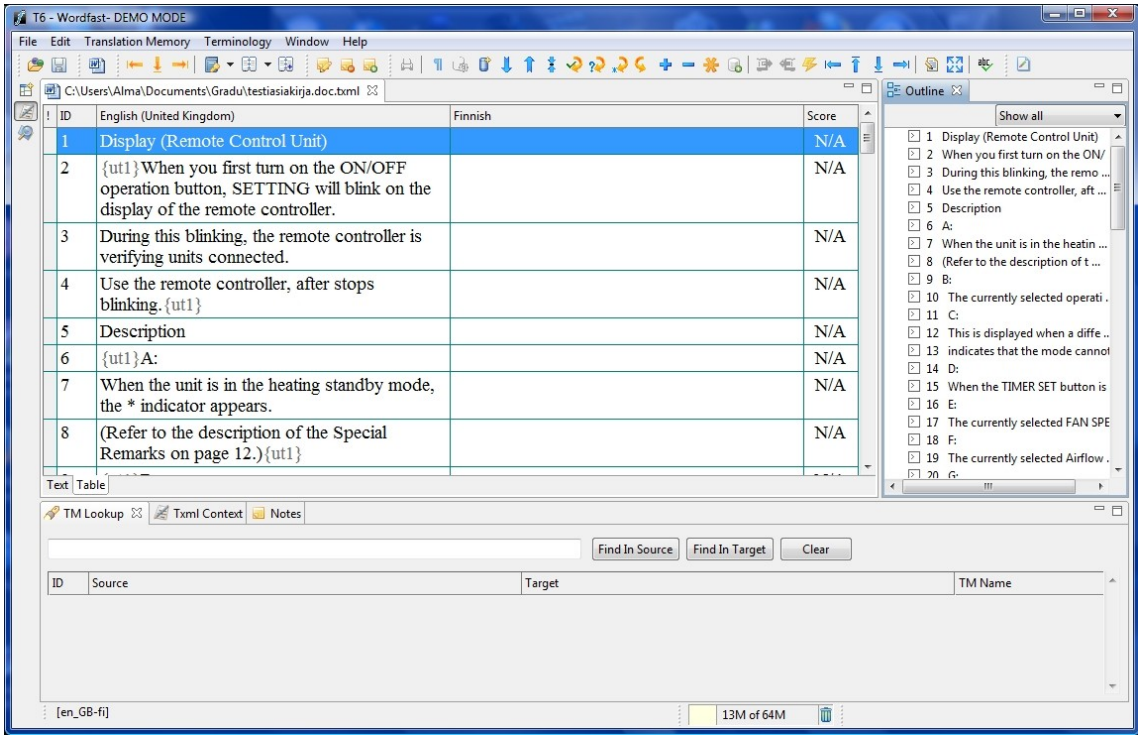


Figure 2. Wordfast Professional interface.

The interface is in English, with no localized versions available.

5. Analysis and discussion

In this chapter, the findings from the usability test are analysed and described. The post-test questionnaires are handled in chapter 5.1, and the screen recordings in 5.2. The material in the questionnaires consists of open questions, therefore there is little quantitative data. The analysis is mainly qualitative; as there is little research about usability in translation technology, the aim is to form an understanding of what constitutes to the usability issues as far as translation students are concerned. Hopefully with the results from this study and other similar studies, in the future the focus can shift to more detailed questions about specific parts of the software.

It should also be noted that not all features of Wordfast Pro were tested; only the core features for translating a text and getting the system running were tested. This calls for further investigation into all features of the software, and a more extensive research

regarding the particular attributes and which parts of the software are difficult for users. To preserve the anonymity of the participants, all of them are referred to as a "she" in the analysis.

5.1. Test results - questionnaires

The following section presents the findings from the questionnaires that participants completed after the usability test. Each participant answered a set of questions related to the software and their performance in the translation test. The questions were largely formulated to produce "yes" or "no" type of answers, but with the possibility of open comments. All participants made further comments about the issues at hand.

From the post-test questionnaires, a few issues emerged that seem to be plain usability problems, and the participants named a number of other issues they felt were difficult. Each test was also recorded and these screen recordings provided further assistance in analysing these issues.

The answers from the questionnaires were analysed and then put into categories on the basis of which part of the system they were related to. First, the results from the post-test questionnaires are described in chapter 5.1.and then, in chapter 5.2, an account of the screen recordings is given.

5.1.1. Appearance and layout

This sub-chapter presents the usability issues that are related to the appearance and layout of the software as described by the participants in the questionnaires.

One of the main issues regarding the appearance was that the participants found the layout somewhat confusing. The participants commented that the drop-down menus were complicated and difficult to navigate, and some functions that they were looking

for were difficult to locate. One feature in particular was difficult to find: the clean-up function. Participants spent considerable time looking for it. It was difficult for some of the participants to find the appropriate setting needed to change the fuzziness threshold, for example, and it was not at all clear to them how changing the settings would affect the overall performance of the software, or what was possible to modify at all.

Some of the participants also thought that traditional menus with descriptive commands would have been easier for them to understand than the icons. Many participants felt that the icons were not clear enough and they did not understand the symbols and their functions. They felt the icons were too small and needed to be clearer, and that there were too many of them. This is a considerable usability issue; the icons should be more self-explanatory than they are at the moment.

5.1.2. Functional issues

This sub-chapter presents the usability issues that are related to the way the software works and how the users complete tasks with it.

The difference between creating a new project and a new translation memory was vague for some of the participants, to the extent that after creating a new project some forgot or did not realise they had to create a new translation memory as well. This resulted in three participants failing to create a new translation memory and only realising their mistake after finishing the translation. Two of these participants created and updated a translation memory afterwards, evidently based on their previous knowledge of translation memory tools.

It certainly has more to do with getting to know translation memories better, but as most of the testers were not very well acquainted with translation memories, and new to this particular version of Wordfast as well, they were unsure what it was that they were supposed to do to complete the tasks. Three of them stated they were not sure what the

different parts of the software were and what it was that they were supposed to do. They were not sure what the standard procedure for translating was, and the software did not offer them any help on this - there is a "my first translation" section on the Wordfast web page, which did not come up through the online manual but had to be searched for separately, as well as some video material, but that had to be searched for separately as well. The participants had, however, used a translation memory tool at least once before as they all had used them for university courses and therefore they knew the principles behind them, which implies that the whole translation memory tool base is overly complicated and difficult to learn. Knowing the basics should facilitate an easy learning phase.

Saving a translated file and performing the clean-up function proved to be difficult. The clean-up command is slightly different in the tested version of Wordfast (Pro) than in the other versions. This task was the point where participants looked at the online help the most. One participant did not even realise she had successfully cleaned and saved the project. She commented on it being illogical and difficult to locate. Also, the difference between the regular save file-command and the clean-up function was not clear to most of the participants. Five out of six participants were not sure whether they had successfully cleaned the file or not. More about cleaning the files in 5.2.3.

Expanding the segments was something a few of the participants would have liked to do, but felt that it was too time-consuming to try to find out how it is done during the test situation, and in a few places Wordfast had divided the text into segments that did not work very well. In one case, Wordfast had divided the source text into segments and when the participant tried to combine these segments, she got an error message saying it could not be done because of a paragraph break. This may have consequences on the translation: to acquire "a natural feel" or to meet some other goal or criteria for the translation a translator may need to combine sentences from separate paragraphs, but the software does not allow it. It may result in forced translations where the translator has to adapt to the situation. The solution is to modify the source text before translating, but the translation tool should support this function and it should be readily available in

an easy-to-use format.

Wordfast has some key combinations for quick access, but the participants were not aware of most of these. Three of the participants used the alt+down command to move from one segment to another. The rest used their mouse and clicked the next segment open but found using just the mouse somewhat tedious. Using the mouse may also make the translation process slower as the user moves from keyboard to mouse and vice versa. A tutorial in all the features of the software would have helped the participants with this. The key combinations did appear when hovering the mouse pointer on the symbols for a few seconds, but it appears some the participants failed to notice this or chose to use the mouse instead.

Two of the participants were unsure whether the translation memory was functioning as it should, because they were not getting many, if any, matches. A point in future studies could be to lower the fuzziness threshold setting of the translation memory: in the present study the participants were asked to set it to 65, but setting it even lower would produce more results and make the participants feel more confident as well as be able to see if the software is performing as it should. Another issue is the quality of the source text - in the current study it was an excerpt from an air conditioner user manual, which contained some repetition. A text with more repetition would naturally provide more matches from the translation memory and thus make it easier for the participants to see if everything is working as it should. Of course, this applies to somewhat inexperienced participants - translators with more experience in using translation tools would no doubt be better equipped to tell if something is going wrong, but at the same time, similarly inexperienced users may buy TM software and try to use it in their work and face similar issues. A symbol indicating the correct usage of the features somewhere on the layout of the software could perhaps work as well; a status update of sorts - for example a red symbol when the feature is not in use, a yellow one if there is a problem with it, and a green one when it is in use and functioning as it should. This way the user could easily check what the status of different features is.

5.1.3. Issues with online help

This chapter deals with issues the participants experienced when reading and browsing the online help section.

It was not always clear in which section the topic they were looking for was posted, and this caused confusion and loss of time, as participants looked for the topics manually and became frustrated. Two participants thought that the manual was not very logically organised.

As pointed out in chapter 5.1.2, the online manual does not include much advice for translators who are unfamiliar with the software. Inexperience caused some confusion among the participants as they did not know what they were supposed to do to complete the tasks. This leads to the conclusion that the online manual should include a section devoted to novice translators, with tutorials and help videos, or the software should include a "quick start/start-up wizard" akin to some Windows software that give the user hints and advice on what to do. At the moment, some tutorials do exist but they are located on a separate internet page and have to be searched for separately. It is not very useful that the help tutorials and videos are scattered on different web sites. Moreover, one participant who had viewed some Wordfast videos before commented that the videos are so tedious and boring they are hard to watch, and the given instructions confusing.

5.1.4. Positive feedback

Based on their previous experiences, all of the participants compared the version of Wordfast used in this test to another version of Wordfast or to another translation memory tool that they had used before and most of them thought that this version, Pro, is quite pleasant to use in terms of layout compared to the other ones they had used. The layout of the TXML Editor perspective in Wordfast Pro was clear and easy for the eye

according to four participants. All of them had used another version of Wordfast, one which integrates with a word processor, and therefore the Pro version was new and different to them. The participants thought it was easier to use.

5.2. Test results - screen recordings

Each performance was recorded with the participant's permission using HyperCam, a screen recording software. The screen recordings were carefully examined, taking into consideration the answers given in the post-test questionnaires as well as information provided in the background questionnaire. These recordings give further insight into the answers given in the questionnaires and provide an excellent view of the issues the participants faced during the test.

Possibly the most useful aspect is to compare the total times in completing tasks during the test. The tasks for which time was measured were
1) how long it took to configure the system, i.e. to create a new project and a translation memory and to adjust the settings as instructed;
2) how long it took the translate the text;
3) how long it took to perform file clean-up and saving, and finally
4) how long each participant spent reading and browsing the online manual.
Personal factors, such as experience in using translation memory tools, naturally affect these times, and this has been taken into consideration when comparing the times. Each category is summed up in the figures below. For the purpose of this study the participants are numbered 1 to 6 as the test was completed anonymously.

5.2.1. Setting up

The following figure presents the time the participants needed to configure the software, i.e. to create a new translation memory, to create a new project, and to adjust the fuzziness threshold as instructed and other settings as they will.

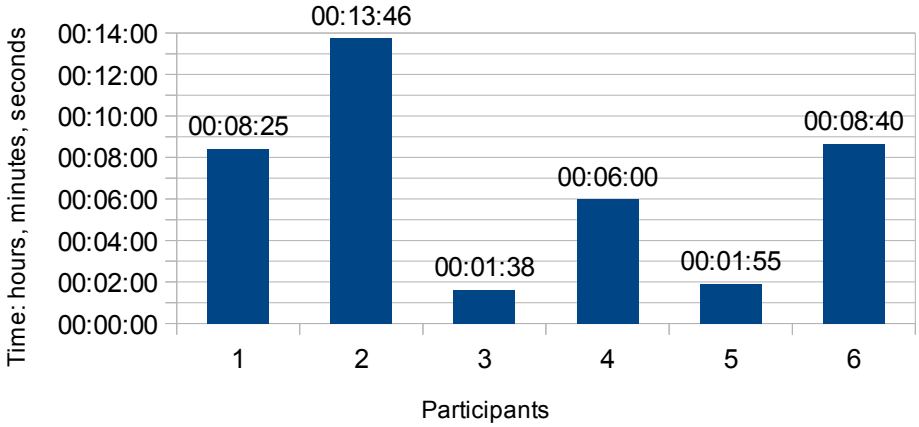


Figure 3. Total time for configuring the system

The average time the participants completed this task in was 00:6:50. Comparing the total times in figure 3 and the corresponding personal details collected before the test in the background questionnaire, it becomes obvious that experience in using translation memories or translation work experience do not seem to play a major role in this task. The fast completion times of participants 3 and 5 (00:01:38 and 00:01:55, respectively) may be in part explained by the fact that they actually did not create a translation memory in the beginning at all, as was instructed. The second fastest participant, no. 5, had experience in translation work and was quite familiar with translation memories, but nonetheless failed to create the translation memory in the beginning as well, instead creating it after finishing the translation and then updating it, as did participant 3. No. 3, the participant with the shortest completion time of 00:01:38, had no work experience in translation, but the participant with the longest set-up time (participant 2 at 00:13:46) had worked in translation. The distinction between creating a new project and a new translation memory seemed to confuse the participants the most, as evident from the questionnaire results as well; they did not fully understand what the difference was, or assumed that creating a new project would automatically create a joint translation memory as well. This is a usability issue that will disappear after some time using the software and getting to know it better, but nevertheless it is an issue for users who are unfamiliar with the system - it caused confusion for three out of six participants.

5.2.2. Translation

Figure 4 shows the difference in time the users finished the translation.

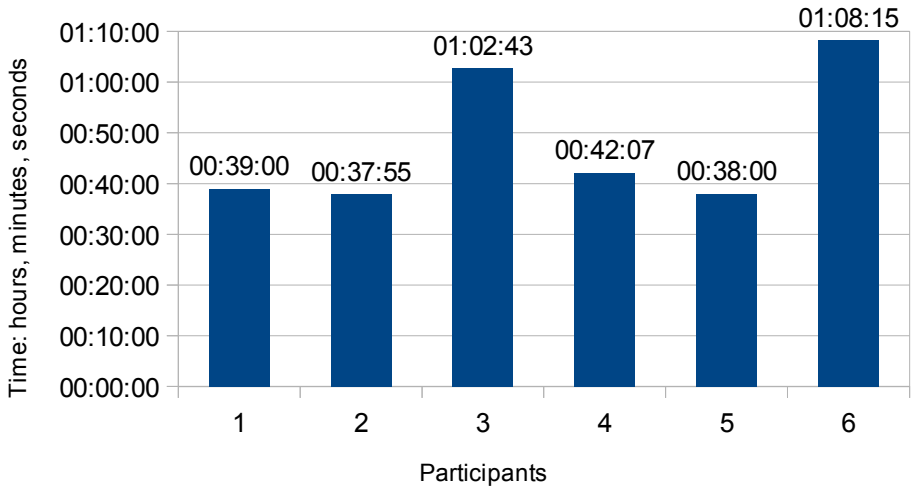


Figure 4. Total time for translation

The average time spent translating is 01:01:20, but this number is slightly skewed by participant 6 because in the process of translating, Wordfast inexplicably crashed and forced the participant to start over as all data completed so far was lost (see below for more). If this is taken into consideration the actual time for translation (after the crash) for no. 6 is 00:34:55, bringing the average down to around 49 minutes. But as is evident, four participants completed the translation in under 40 minutes, if no. 6 is counted with the revised completion time. As such, the time spent translating is not an easily comparable entity in that there are many variables that contribute to it, be it personal preferences, circumstances or other variables such as level of ambition with a test translation. The outcome will vary greatly depending on these individual factors. In the scope of this study this was overlooked for reasons of simplicity in the hopes that the overall time reflects the ease of use that participants experience; however direct conclusions cannot be drawn based on this. An obvious factor affecting participant 6's work was the unexpected system crash. Others experienced other minor issues and

inconveniences, and on the whole the translation process seemed to flow quite naturally for all participants. Some issues were raised which have previously been mentioned about the quick access key combinations and the lack of instructions in the online manual.

Two of the participants, who were both among the participants who did not create a translation memory in the beginning of the task, encountered the following error message several times during the translation process: "No active TM server. Create new or select an existing one", which both of them ignored and continued translating. This error message was not included in the troubleshooting section and there was no mention of it in the online manual, but searching on the internet by the present author provided some results: it seems this error message appears when a user has not ticked the "selected" box for the translation memory when on the "New/Select TM" tab, or has ticked the "read-only" option. The same appears to be the solution for another error message both of these same participants encountered: "Commit to TM: There is no writable translation memory selected". This error message was not found in the manual either, but the present author searched for it on the internet. Not describing all of the error messages is an obvious flaw in the user manual. Another interesting aspect here is the behaviour of the participants: both participants who encountered these error messages viewed them, clicked "OK" and then continued to work on their translation as if nothing had happened. It could be an over-saturation of error messages that current technology has brought upon users, or some other reason for ignoring these messages, but nevertheless the participants did not seem bothered about these error messages, although they should have. One of them mentioned in the questionnaire that she did feel rather silly for not noticing that they had not selected a translation memory.

Participant 6 encountered a fatal error during the translation process: halfway through the translation the software inexplicably crashed. All data completed so far was lost and the participant had to start from the beginning. It took nearly seven minutes for the system and user to recover. An error message was produced after using the Task Manager to close Wordfast, which did not respond, which said: JAVA (TM) Platform.

"SE binary is not responding." A long crash report was produced with the following information in the beginning: "JVM terminated. Exit code = -805306369". This appears to be some sort of Windows Vista problem, according to searches on the internet done by the present author, but more information about this specific error could not be found, and the reason for the crash remained a mystery.

Another error in Wordfast encountered by one participant was the following: "TXML file is missing. Save -- as TXML and try again." Again, this error was not found in the manual. All of the error messages should be explained in detail in the manual, with solutions provided. Also, when looking into the matter further, there is no explanation what a TXML file actually is.

Only one participant used the spell-check in Wordfast. The spell-check did not provide any useful corrections: all of the "corrections" it offered were irrational, and therefore not useful at all. Most of them were misspelled or completely different words, for example for "pysäyttäminen" Wordfast suggested "pystyttäminen", and most of the suggestions it made were compound nouns written separately that were originally correct in the text. Also, one participant chose to tick "copy source on no match" in the preferences, and then had to manually omit the source language sentences from the target segment side, which seems like extra work one could do without.

An issue with this test in particular was that a number of participants commented that they had not looked into for example changing some setting because they felt it would be too time-consuming in the scope of the test situation to try to learn how that particular feature or setting works - they seemed to lack motivation in a testing situation like this. This diminishes the possibilities of discovering potential usability issues and test participants should be encouraged to research all areas they feel necessary and to complete tasks as if they were real assignments. It is quite understandable from a human perspective that the participants did not want to spend too long with this one assignment. Completing the test should therefore be made more appealing for the testers, and quite a lot of usability testers do offer rewards for participants. A nice

reward at the end would encourage participation - or maybe it would encourage the users to finish the assignments as quickly as they can so they can get the reward, leading to the same problem. Perhaps participants should be given a small reward after each task - this way encouraging completing the tasks more thoroughly. This is, however, a very primitive view - that the only motivation for people is a reward. During the test situation it did seem like participants were impatient to finish and did not want to spend time learning extra features that may have helped them with their tasks because of the added time, or did not spend time for example watching tutorial videos.

The conclusion is that the user manual needs a more extensive troubleshooting section with potential error messages, their origins and solutions presented in an understandable way. In future studies, the reward policy should be considered.

5.2.3. Saving and cleaning

Figure 5 presents the total time each participant spent cleaning and saving the file.

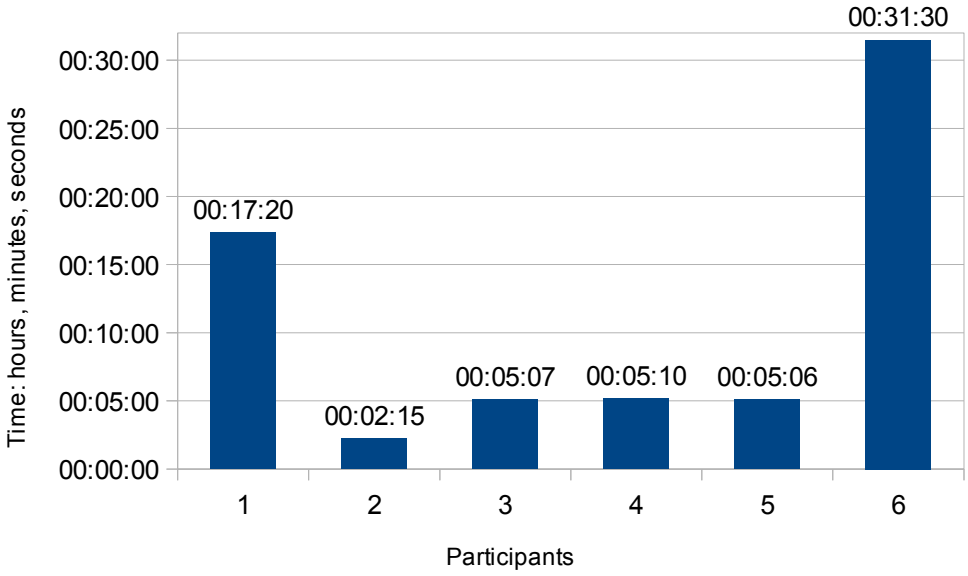


Figure 5. Total time for clean-up and saving

The average time here is 00:10:48. Based on the figures in figure 5, particularly one

participant found it difficult to clean and save the file. However, the clean-up function in this version of Wordfast is very simple and straightforward: all the user needs to do is to select the "Save Translated File" command. All of the participants had used a different translation memory tool in the past and therefore expected the clean-up function to be different, similar to Wordfast Classic; a few participants expressed that they were expecting something confusing and difficult to use based on their past experiences using the version of Wordfast which is integrated in a word processing software. The instructions for the test told the participants to "clean and save the file", which may have caused part of the confusion in this task as the "Save Translated File" command does not mention cleaning and the participants were specifically looking for a clean-up command. The instructions had to be worded that way to avoid confusing the regular save command with the "Save Translated File" command. Otherwise it may have not been clear to the participants that they were expected to clean the file and, furthermore, it would have likely marred the point of the study to plainly tell the participants to use the "Save Translated File" command, when the point is to test how easy it is to learn how to use and then use the software. All of the above, combined with the preliminary expectations of the participants, led to the confusion in this task. Admittedly, the manual does not explain the actual function of the command very clearly, and only notes that it can be used "...to save a translated file to its original format." As it was, nearly all of the participants, particularly no. 6, spent a lot of time trying to locate a separate clean-up icon or function in the software, and eventually finding and using the clean-up function in the project manager perspective of Wordfast, which also does perform the clean-up function but is more complicated and more difficult to locate. Moreover, this feature is usually used to clean batches of files and for updating translation memories. All of the participants successfully cleaned the file in the end.

All in all, this step proves that careful wording when instructing is important and that preliminary expectations can lead to confusion as one expects the software to work according to previously learned patterns.

5.2.4. Online manual

Figure 6 presents the total time each participant spent reading and browsing the Wordfast online manual accessed through the Help button in the Help drop-down menu.

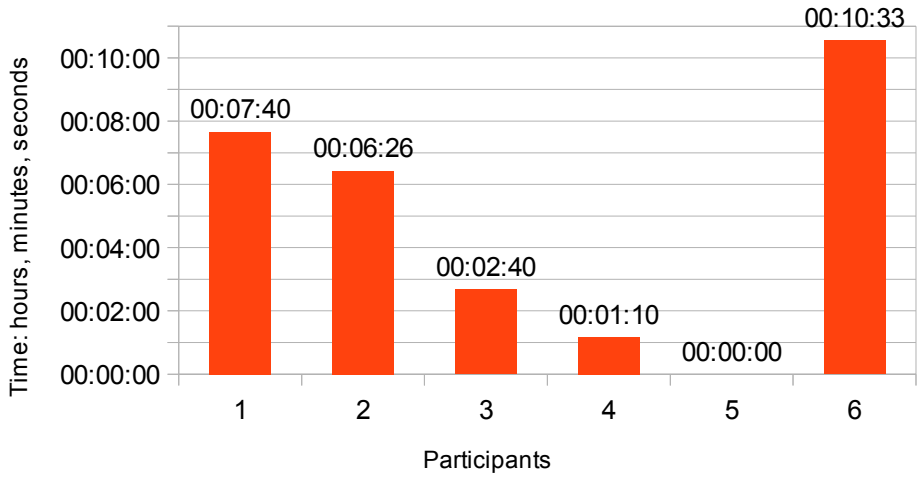


Figure 6. Total time reading manual.

The average time a participant spent reading the manual was 00:04:58. One could hypothesise that the longer a person spends reading the manual, the more difficult they find using the software, or the less they have used translation memories before. In this case every participant was familiar with translation memories and how they work; that should eliminate some of the time attributed to not knowing what the system does. In this task, the amount of experience with translation memories as described by the participants themselves on the background questionnaire seems to roughly correspond with the time they spent reading the manual. The participants with less experience spent more time reading the manual. Notably, one participant did not look at the manual at all, based on their previous experience in using translation memories as stated in the background questionnaire and also in the answers in the post-test questionnaire. The total time gives an indication as to who is or considers themselves to be a confident user. Even the person who did not look at the manual made some errors in the test, which were, however, probably caused by not reading the instructions carefully enough. Moreover, this seemed to be the tendency among nearly all of the participants: they skimmed over the instructions and started translating, whereas if they had carefully read

the instructions they might have avoided the issue of not creating a translation memory in the beginning of the task, which three of them neglected to do. They were specifically and in detail told in the instructions to do so. It is therefore crucial to write instructions that are clear, short and precise: it appears participants do not necessarily spend time reading them very carefully.

5.3 Overview and categorisation of usability issues

This chapter will give an overview of the usability issues that were uncovered during the usability test process and described in chapters 5.1 and 5.2. Usability issues were categorised according to the following scale presented in chapter 3.7.5:
0 - not really a usability issue, just differences in opinion by the examiners
1 - cosmetic usability issue, will be fixed when and if there's time and money.
2 - minor usability issue. Low priority, will be fixed after more serious issues have been fixed.
3 - Serious usability issue, high priority, should be fixed as soon as possible.
4 - Catastrophic usability issue, prevents the intended use of the application, must be fixed immediately.

As the author of this study is by no means experienced in usability testing or a proficient translation memory user, the forming of serious suggestions is omitted and left for more qualified researchers to tackle. However, some thoughts and ideas regarding the matters are presented. Hopefully this study can, nevertheless, draw some attention to the requirements of translators as technology users.

All of the improvement suggestions presented in this study are made by the present author and have not been inspected or approved by a usability specialist. The following categorisation is based on the opinion of the present author only and is therefore subject to some reservations. The ranks are awarded from the point of view of a beginning translator, such as the participants in this study, and therefore easy learnability is

stressed.

The main issues with the layout that the participants pointed out were: 1) the icons and symbols were not self-explanatory enough; 2) the icons needed to be larger and clearer; and 3) that the menus were confusing and difficult to navigate. The first and the second issue with the confusing icons are classed as level 2, because the unreadability of the icons to persons who are already familiar with the principles of translation memories does form a usability issue, but not one that would hinder the intended use of the program severely - it merely slows down the process. The third issue with the layout, the difficult menus, is classed as level 2 as well for similar reasons. Initial difficulties in navigating the menus and the icons prove that their functions are not self-evident.

Out of the functional issues the most problematic were: 1) the confusion with creating new projects and new translation memories, and 2) the procedure for cleaning translated files. The confusion participants experienced with creating projects and translation memories when starting their test translation is a major problem and does prevent the intended use of the program, the use of the translation memory, and it is therefore classed as level 4. As is evident from chapter 5.2, three of the six participants failed to create a translation memory when they were setting up the system, only realising their mistake after finishing the translation. The function for updating the translation memory after translation does not remove the problem here: the intended use of the software is to get matches from the translation memory while translating, and that certainly does not happen if there is no translation memory selected. An easy solution presents itself: incorporating the selection or creation of a translation memory into the process of creating a new project. The first step would be to create a new project, the second to create a new translation memory for the project or select an existing one, and finally the third to adjust the settings as desired. If all these functions were included in the same stage of creating a new project, and automatically came up after one another, the process of starting a new translation would seem less confusing. A similar idea to facilitate the intended use of the software would be the coloured symbols mentioned in chapter 5.1.2 that would indicate that all parts of the system are running as they should,

or if there was an error, or if they were not in use at all. The second problem with cleaning files is a minor one compared to the issue with creating projects as it does not obstruct the intended use of the software, however it did result in some confusion. It is classed as level 2 because the function should be more self-explanatory.

The difficulties in expanding segments and merging segments from two paragraphs is a fairly serious issue but does not prevent the intended usage of the software; therefore it is classed as level 3. As mentioned in chapter 4.1.2, not being able to freely combine segments can affect the end translation in an undesired way. The only way around this issue seems to be to edit the source text.

The confusion over the overall translation process is a minor issue, which is classed as level 2, and which an increase in familiarity with Wordfast should fix. Nevertheless it would benefit new users very much to have an integrated start-up wizard or tutorial of some sort to help them get through the difficult first phase. This issue is connected to the overall larger issue of help documentation.

The issues with the online manual are slightly more severe and should be remedied as soon as possible, and are therefore classed as level 4 because of their immediate effects on the use of the software. The lack of advice for beginners in the attached help section is a major problem, as is the lack of a complete troubleshooting section and the lack of explanations for the error messages. The difficulties some of the participants experienced when navigating the help manual are minor compared to the larger issues with the help manual, but still should be classed as level 2 for their impact on the usability of the manual.

Another issue which is classed as level 2 is the spell-check, which offered odd, misspelled matches to the participant who used it. The language settings seemed to be right, but it nonetheless offered words with completely different meanings as alternatives.

5.4. Translation technology training

This section is handled separately as it is unconnected to the rest of the study as such. In the same post-test questionnaire in which the participants were asked questions about the usability issues, they were also asked to give their thoughts on the state of translation technology teaching they have received. All of the participants were students enrolled at the University of Eastern Finland, with English and translation as their major, and had therefore taken part on roughly the same courses. They were asked how much teaching in translation tools and technology they had received, whether they thought it was sufficient, whether they thought it had prepared them well for the demands of working life, and whether they thought that more courses should or should not be taught in translation technology. The assumption was, based on observations attending the University of Easter Finland, that most participants would like to learn more about translation tools and use them more on courses. This assumption was supported by the participants' responses.

The prevalent opinion among the participants was that translation tools are not in a prominent enough position in the curriculum of translation studies. All had taken part in at least one course that deals with translation technology, but felt that it had not given them the skills they felt are required by employers. As one participant put it: "The course barely scratches the surface, and uses the worst version of Wordfast coupled with poor instructions" (translations by the present author). They commented that the importance of translation tools is underestimated. A few of the participants severely questioned the university's ability to teach them the necessary skills for working life, such as this participant, when asked if she thought the courses had prepared her well for the technological requirements of working life: "Certainly not, but this study programme doesn't prepare you for any other aspects of working life either." The message was clear: more teaching in translation technology and a more hands-on approach to translation. A few comments were made about the bias on translation tools in teaching: for example it seems corpora have received more attention than other tools, and when considering their usefulness to a professional translator the participants felt

this was not justified. Every participant thought that translation tools should be used more on translation courses to familiarise students with them. As pointed out in chapter 2.4, the COLC project at the Savonlinna School of Translation Studies, which introduced a lot of translation technology into the curriculum, was highly successful, and it seems a similar approach is needed. A working knowledge of translation tools is required by most employers today, after all.

To sum up, every participant felt strongly that more courses on translation technology should be offered and that electronic translation tools should be used on many more courses. None of the participants thought that the courses offered at university had prepared them very well for the technical demands of the profession.

5.5 Methodological evaluation

This chapter gives an account of how this study was conducted, what could have been done differently and what caused problems during testing to facilitate future testing. One of the most important things to do is to conduct a pilot study to test the test procedures and the test materials. It may also be necessary to retest materials that have been altered. That was the case in the current study - the first version of the post-test questionnaire was rejected and a new version drawn, but only the first version was tested. The new version of the questionnaire should have been subjected to testing as well to make sure it was not ambiguous and that the questions provided desired results. As it was, some of the questions did not produce quite the answers that were hoped for. Careful wording is critical in questionnaires to avoid leaving room for interpretation.

Careful wording and careful thinking is also necessary when creating the background questionnaire and other necessary documents. There were a few things that would have provided valuable information to the current study which were not asked in the background questionnaire; for example year of studies. This is also another example of wording: the participants were asked for their level of education - but not for their year

of study. There is a clear difference there that should have been identified before actual testing. Other test materials need to be carefully considered as well: for example the source text for translation is in a critical role in a study like this, particularly because with the right kind of text, participants should see frequent matches from the translation memory. The text chosen for this test was not ideal, after all, even though it seemed like it in the beginning, and coupled with the fuzziness threshold that was perhaps set too high, the participants did not get many matches from the TM. To prevent this it is advisable to choose a text that is relatively simple and that contains many repetitious elements, and to set the fuzziness lower than one would in a normal translation situation, as mentioned in chapter 5.1.2. It could possibly alleviate some of the frustration the participants experienced during the test when they were unsure if the translation memory was working or not.

Recruiting participants can be challenging - testers looking to recruit should exploit multiple forums to find as many as possible, depending on their study of course. The recruitment process was quite difficult for this study as well, and only six participants volunteered. All of the participants came from a similar academic background as well, meaning that there was no diversity, but on the other hand at least they form a sample base where it is easier to compare one against another. In future studies it would be very interesting to have experienced translators take part in similar tests and answer similar questions about usability, and also to compare their answers against a pool of answers from translator students. Even just a questionnaire without any actual usability testing could be quite easily done via e-mail and internet, reaching a wide base of professionals in the translation industry. A reward scheme should also be considered to attract more participants.

Another feature that could be monitored is key strokes, which would reveal more about how the participants go through the translation and which key combinations they use. The screen recordings taken in this study proved to be very useful, but they are very time-consuming to analyse, therefore screen recordings are most likely not the best choice in studies with many participants unless there is a lot of time and resources

available.

An issue to consider is that this study was made after a finalised product has already been released. Ideally usability testing takes place before publication. Nevertheless, this study will hopefully be the first of many to point out the issues in translation tools. A framework for evaluating translation tools is needed, as suggested in chapter 3.3, and hopefully this can be the starting point for future studies.

As pointed out before, the author of this study is not a usability expert and the views expressed in this study may not be what usability experts would stress, but hopefully this study will be of some value in drawing attention to the issues of usability in the translation tools industry. It has at least shown that the learning phase for Wordfast Pro is not very easy for translation students.

6. Conclusion

This study set out to explore the difficulties faced by translators when using translation memory tools, in this particular case Wordfast Pro. The usability test conducted proved that there are some usability issues that at least starting translators face when using the software, particularly when using it for the first time.

The study began with a review of relevant background theory from translation studies and usability research. Based on this, a usability test was planned and conducted to reveal the particular problems that users encountered when translating a short sample translation. The findings from the usability test were carefully analysed and described. In the test, participants were asked to translate an excerpt from a user manual using Wordfast Pro, and then answer questions about the software and their performance in a questionnaire. Each test was also recorded with a screen recording software to further examine the process.

The results of this study are indicative only - further research into the causes and implications of usability in translation tools is required. Due to the small number of test participants statistically valid results could not be obtained, but nonetheless the results of this study point to some more serious and some less serious usability problems that could be discovered even with an informal study and with few participants. What constitutes to a usability problem can be defined differently by different test organisers, and the results in this study only reflect the view and understanding of the current author.

It is also important to note that there are several factors that affect the outcome of this study; the number of participants, their personal attributes, the limited resources available etc. The purpose of the study has in any case been fulfilled: to uncover as many usability issues as possible with the material available and to form a better understanding of what causes these issues. These usability issues were uncovered using a combined method of questionnaires and screen captures. Both were examined carefully taking into consideration the personal attributes of each participant. The found usability issues were categorised according to their seriousness.

Creating a new translation memory proved to pose the most problems to the participants in the context of starting a new translation and creating a new project. The distinction between creating a new project and creating a new translation memory was not obvious enough. Another similar issue was the ambiguity concerning file clean-up: the command for cleaning a translated file was difficult for participants to identify and find, and the user manual did not explain it carefully enough.

The layout of Wordfast received some criticism: many of the participants thought that the icons were too small and difficult, sometimes impossible, to interpret. The drop-down menus were difficult to navigate and confusing, and it was difficult to find the appropriate settings.

The help manual was another source of discontent for the participants. They felt the

manual was poorly structured, confusing and difficult to navigate. The manual did not include a section devoted to starting translators, or a complete troubleshooting section. Tutorials for new users were scattered across other internet pages, and none were on the manual itself. Other issues include difficulties in merging segments, and a spell-check tool that did not quite work the way it should.

Overall, the problems faced by the participants can be partly attributed to inexperience. However it must be stressed that all of the participants had used a translation memory tool previously, and were not complete beginners. Taking this into consideration it seems the translation process here should have been relatively easy. Learnability and memorability are key aspects of usability, after all. This can only lead to the conclusion that Wordfast Pro is not a user-friendly tool in these aspects.

Based on the experiences in this study, it seems more usability research is indeed needed in the translation industry. More detailed tests need to be conducted with larger and more varied user groups to make valid claims about usability issues. The context of translation is an important factor to consider, too: what restrictions or stipulations it poses on the software itself and also on the testing process.

REFERENCES

Austermühl, F. 2001. *Electronic tools for translators*. Manchester: St. Jerome Publishing.

Baker, M. 1993. Corpus Linguistics and Translation Studies: Implications and Applications in Baker, M, Gill, F. and Tognini-Bonelli, E. (eds.) *Text and technology: in honour of John Sinclair*. Philadelphia: John Benjamins Publishing Company.

Bowker, L. 2002. *Computer-aided translation technology: a practical introduction*. Ottawa: University of Ottawa Press.

Bowker, L. and Fisher, D. 2010. Computer-aided translation in Gambier, Y. and Van Doorslaer, L. (eds.) *Handbook of translation studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Dragsted, Barbara. 2004. *Segmentation in translation and translation memory systems. An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process*. Copenhagen: Copenhagen Business School.

Dumas, J. S. and Redish, J. C. 1999. *A practical guide to usability testing*. *Revised Edition*. Exeter: Intellect Books.

Esselink, B. 2000. *A practical guide to localization*. Amsterdam: John Benjamins B.V.

Granger, S. 2003. "The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies?" In Granger, L. and Petch-Tyson, S. (eds.): *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi B.V.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Höge, M. 2002. *Towards a Framework for the Evaluation of Translators' Aids' Systems*. Helsinki: Helsinki: University of Helsinki.

IEEE 1012-2004 *Standard for Software Verification and Validation*. Institute of Electrical and Electronics Engineers, 2005.

Jaatinen, H. and Jääskeläinen, R. 2006. Introducing IT in translator training: Experiences from the COLC project in Pym, A. Perekrestenko, A. and Starink, B. (eds.) 2006. *Translation technology and its teaching: with much mention of localization*. Tarragona: Universitat Rovira i Virgili.

Jääskeläinen, R. and Mauranen, A. 2000. *Project SPIRIT. Work Package 5: Development of a Corpus on the Timber Industry, Final Report.* University of Joensuu.

Jääskeläinen, R. and Mauranen, A. 2001. Kääntäjät ja kieliteknologia - kokemuksia työelämästä. in Charles, M. and Hiidenmaa, P. (eds.): *Tietotyön yhteiskunta - kielen valtakunta. AFinLAn vuosikirja 2001. Suomen soveltavan kielitieteen yhdistyksen julkaisuja nro 59.* Jyväskylä.

Kenny, D. 2011. Electronic tools and resources for translators in Malmkjaer, K. and Windle, K. (eds.) *The Oxford handbook of translation studies.* Oxford: Oxford University Press.

Kitchenham, B. A. and Pfleeger, S. L. 2002. Principles of Survey Research Part 3: Constructing a Survey Instrument in *Software Engineering Notes* vol 27 no 2. ACM SIGSOFT.

Koskinen, J. 2005. Käytettävyystutkimus in Ovaska, S. Aula, A. and Majaranta, P. (eds).

Käytettävyystutkimuksen menetelmät. Tampere: Tampereen Yliopisto.

Laviosa, S. Corpora and Translation Studies. In Granger, L. and Petch-Tyson, S.(eds.) 2003. *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi B.V.

Nielsen, J. 1993. *Usability Engineering*. San Diego: Academic Press.

Olohan, M. 2004. *Introducing Corpora in Translation Studies*. London: Routledge.

Ovaska, S. Aula, A. and Majaranta, P. (eds). 2005 *Käytettävyystutkimuksen menetelmät*. Tampere: University of Tampere.

Pym, A. Perekrestenko, A. and Starink, B. (eds.) 2006. *Translation technology and its teaching: with much mention of localization*. Tarragona: Universitat Rovira i Virgili.

Quah, C.K. 2006. *Translation and technology*. Basingstoke: Palgrave Macmillan.

Ravden, S. and Johnson, G. 1989. *Evaluating usability of human-computer interfaces: a practical method.* Chichester: Ellis Horwood Limited.

Rubin, J. 1994. *Handbook of Usability Testing. How to plan, design, and conduct effective tests.* New York: John Wiley & Sons.

Salmi, L. 2002. Computers, Documentation and Localisation: A Methodological Perspective in *Across Languages and Cultures. A Multidisciplinary Journal for Translation and Interpreting Studies*. Volume 3, Issue 1. Budapest: Akadémiai Kiadó.

Sharp, H, Rogers, Y. and Preece, J. 2007. *Interaction design. Beyond human-computer interaction.* Chichester: John Wiley & Sons Ltd.

Somers, Harold (ed.) 2003. *Computers and Translation. A translator's guide*. Amsterdam: John Benjamins Publishing Company.

Vanhala, T. 2005. Kyselylomakkeet käytettävyystutkimuksessa in Ovaska, S. Aula, A. and Majaranta, P. (eds) *Käytettävyystutkimuksen menetelmät*. Tampere: Tampereen Yliopisto.

White, J. S. 2003. How to evaluate machine translation in Somers, Harold (ed.). *Computers and Translation. A translator's guide*. Amsterdam: John Benjamins Publishing Company.

Wixon, D. and Wilson, C. 1997. The usability engineering framework for product design and evaluation. In Helander, M.G, Landauer T.K. and Prabju P.V. (eds): *Handbook of Human-computer Interaction*. Amsterdam: Elsevier.

Online sources:

EAGLES
http://www.ilc.cnr.it/EAGLES/home.html
http://www.issco.unige.ch/en/research/projects/ewg96/node2.html#SECTION00200000000000000000
accessed 25.10.2009

European Association for Machine Translation
http://www.eamt.org/mt.php, accessed 8.10.2009

Cone Tree
http://www.conetrees.com/2010/12/ux-glossary/post-study-system-usability-questionnaire-pssuq/
accessed 18.10.2011

LISA - Localisation Industry Standards Association
http://www.lisa.org/, accessed 12.01.2010

Sanyo
http://us.sanyo.com/dynamic/product/Downloads/Install_and_User_PAC_72_Series_Remote_Controllers-5728016.pdf, accessed 12.12.2011

FINNISH ABSTRACT

Tutkielman aiheena on tutkia käännösmuistiohjelmien käytettävyyttä ohjelman käyttäjän näkökulmasta. Käytettävyystutkimusta on tehty kääntäjien sähköisten työkalujen kanssa vain vähän, ja materiaalia siitä on saatavilla erittäin rajoitetusti. Vaikuttaakin enemmän siltä, että ohjelmat on suunniteltu ottamatta juuri huomioon kohdekäyttäjien toiveita tai tarpeita (Jääskeläinen ja Mauranen 2001).

Tutkielman ensimmäinen osa keskittyy kuvaamaan tutkimukseen liittyvää teoriataustaa sekä käännöstieteen että käytettävyystutkimuksen näkökulmista. Erityisesti käytettävyystestauksen eri prosessit kuvataan tarkasti.

Tavoitteena on muodostaa kuva siitä, minkälaiset asiat aiheuttavat käyttäjille ongelmia tai vaikeuksia käännösmuistiohjelmien käytössä ja löytää mahdolliset ongelmat testissä käytetystä ohjelmasta. Tutkimusta varten järjestetyssä käytettävyystestissä vapaaehtoiset osallistujat käänsivät tekstikatkelman lämpöpumpun käyttöohjeesta käyttäen Wordfast Pro -käännösmuistiohjelmaa. Osallistujat vastasivat sen jälkeen kyselyyn, jossa kysyttiin heidän mielipiteitään ohjelman käytettävyydestä ja testin sujumisesta. Lisäksi jokainen testitapahtuma tallennettiin käyttäen kuvaruutunauhoitusohjelmaa kääntämisprosessin lähemmän tarkkailun ja vertailun mahdollistamiseksi.

Pohjaolettamus testille oli, että useimmat kääntäjät kokevat jonkinasteisia ongelmia käyttäessään käännösmuisteja. Olettamus perustuu henkilökohtaisiin kokemuksiin. Testi osoitti olettamuksen todeksi ainakin aloittelevien kääntäjien osalta. Lisätutkimusta aiheesta kaivattaisiin; tähän tutkimukseen ei osallistunut muita kuin opiskelijoita. Muitten käyttäjäryhmien testaaminen toisi erilaisia tuloksia ja uusia ulottuvuuksia testaukseen.

Testin ongelmaksi osoittautui osanottajien vähyys sekä tutkimuksentekijän kokemattomuus käytettävyystutkimuksessa. Ensimmäisen vuoksi tutkimuksen tuloksia

on vaikea yleistää, ja toinen aiheutti joitakin epäselvyyksiä tutkimusprosessissa sekä muutamia aloittelijan virheitä, jotka olisi voitu välttää.

Käytettävyystestin järjestämisessä haasteena oli sopivien osanottajien löytäminen. Loppujen lopuksi kuusi vapaaehtoista suoritti testin. Näin pienellä osanottajamäärällä ei valitettavasti pystytty tuottamaan tilastollisesti päteviä tuloksia, mutta pystyttiin kuitenkin paikallistamaan suurimmat käytettävyysongelmat. Kaikki osanottajat olivat englannin kielen ja kääntämisen opiskelijoita Itä-Suomen yliopistosta, ja kaikki olivat ottaneet osaa kurssille tai kursseille, joilla opetetaan käännösteknologian käyttöä. Käännösmuistien perusperiaatteet olivat siis osanottajilla tiedossa, vaikkei käytännön kokemusta olisi paljon ollutkaan.

Itse testitilanteessa osanottajat täyttivät esitietolomakkeen itsestään, allekirjoittivat nauhoitusluvan, ja tekstin kääntämisen jälkeen täyttivät kyselylomakkeen. Osanottajat kohtasivat eri tyyppisiä vaikeuksia, mutta joitakin selkeitä käytettävyysongelmia pystyttiin rajaamaan. Yksi näistä oli sekaannus uuden projektin luomisvaiheessa: puolet testaajista unohti tai ei huomannut luoda uutta käännösmuistia, vaan käänsi tekstin ilman käännösmuistia, koska ei ollut ymmärtänyt tai huomannut eroa uuden projektin ja uuden käännösmuistin luomisessa. Toinen merkittävä käytettävyysongelma oli ohjelman ulkonäkö: osanottajat pitivät ikoneita liian pieninä ja liian epäselvinä, ja kommentoivat valikkojen sekavuutta ja toimimattomuutta. Kolmas ohjelman käyttöön liittyvä merkittävä ongelma oli käännössegmenttien yhdistäminen, joka ei onnistunut, jos segmenttien välillä oli kappaleenvaihto. Myös käännetyn asiakirjan puhdistustoiminto aiheutti päänvaivaa, sillä osanottajat etsivät vanhasta kokemuksesta erillistä puhdistustoimintoa, vaikka sellaista ei olisi varsinaisesti tarvinnutkaan käyttää. Osanottajat olivat aiemmin käyttäneet eri versiota Wordfastista, jossa puhdistustoiminto toimii eri tavalla, ja sovelsivat tätä vanhaa tietoa testattavaankin versioon.

Muita ongelmia aiheuttivat online help -osion sekavuus ja puuttellisuus. Testaajat eivät helposti löytäneet haluamaansa tietoa sieltä, eikä osiosta löytynyt ratkaisuja esimerkiksi ohjelman antamiin virheviesteihin eikä osiota aloittelevien käyttäjien avuksi. Wordfastin

korjauslukutoiminto ei myöskään toiminut odotetulla tavalla, vaan antoi kummallisia korjausehdotuksia.

Testin tulokset osoittivat alkuolettamuksen oikeaksi: yksikään osanottaja ei selvinnyt testistä ilman jonkinasteisia vaikeuksia tai ongelmia. Testin osallistujat olivat kaikki uusia käyttäjiä; tästä voidaan vetää johtopäätös, että Wordfast Professional ei vastaa käytettävyyden kriteereihin ohjelman helpon oppimisen ja muistettavuuden kannalta.

Testin jälkeen täytetyssä kyselyssä kysyttiin myös osallistujien mielipidettä käännösteknologian opetuksesta. Tässä osiossa vastaukset olivat kaikilla osanottajilla yhteneväisiä: kaikki toivoivat lisää käännösteknologisten työvälineiden opetusta sekä enemmän työvälineiden käyttöä kursseilla. Testaajat eivät kokeneet, että yliopistossa tarjotut kurssit olisivat valmistaneet heitä erityisen hyvin työelämän tarpeisiin, ja että käännösteknologisten työvälineiden merkitystä aliarvioitiin.

Käytettävyystestistä saadut tulokset antavat suuntaa tulevalle tutkimukselle. Lisätutkimus on tarpeen, jossa suurempi määrä osanottajia erilaisista käyttäjäryhmistä otetaan mukaan. Erilaisten käyttäjäryhmien vertailu toisi myös mielenkiintoisia tutkimustuloksia. Myös käännösasiayhteyden merkitystä käytettävyydelle tulisi tutkia: kääntäminen prosessina aiheuttaa kuitenkin jonkinlaisia vaatimuksia ja rajoitteita niin ohjelmistolle kuin käytettävyydellekin. Käytettävyyden merkitystä käännösteknologisisa ohjelmistoissa ei tulisi aliarvoida, ja itse käyttäjät tulisi ottaa mukaan suunnittelu- ja testausprosessiin.

APPENDIX 1. BACKGROUND QUESTIONNAIRE.

Esitietolomake: Wordfast käytettävyystesti.

Sukupuoli:
Ikä:
Koulutustausta ja -taso:
Yleinen kokemus tietotekniikasta (taso, vuodet):
Oletko käyttänyt käännösmuistiohjelmia ennen? Jos olet, kuinka paljon:
Ovatko käännösteknologiset työvälineet ylipäätään tuttuja:
Onko sinulla alan työkokemusta:

APPENDIX 2. POST-TEST QUESTIONNAIRE.

1. Ohjelman layout ja rakenne:
1.1. Onko ohjelman ulkonäkö mielestäsi selkeä ja ymmärrettävä?

1.2.Ymmärrätkö, mitä ohjelman eri osat ovat?

2. Projektin luominen:
2.1. Onnistuiko uuden projektin luominen helposti?

2.2. Oliko asetusten muuttaminen helppoa tai vaikeaa?

2.3. Mitä ongelmia tai vaikeuksia kohtasit projektia luodessa?

3. Käännösmuistin luominen:
3.1. Onnistuiko uuden käännösmuistin luominen helposti?

3.2. Oliko asetusten muuttaminen helppoa tai vaikeaa?

3.3. Mitä ongelmia tai vaikeuksia kohtasit käännösmuistin luomisessa?

4. Kääntämisen aloittaminen:
4.1. Oliko kääntämisen aloittaminen helppoa?

4.2. Mitä ongelmia tai vaikeuksia kohtasit aloittaessasi kääntämistä?

5. Kääntäminen:
5.1.Oliko kääntäminen helppoa käytettävyyttä ajatellen (ei siis esim. terminologia tai muu ajatustyö, vaan ohjelman toiminta)?

5.2 a) Mitä eri toimintoja käytit? b) Oliko eri toimintojen käyttäminen helppoa ja selkeää?

5.3. Oliko sinulle selvää, mitä toimintoja tulee tehdä ja miten ne tehdään?

5.4. Oliko pikanäppäimien käyttö mahdollista, käytitkö niitä?

5.5. Oliko asiakirjan jako segmentteihin selkeä?

5.6. Toimiko käännösmuisti odotetusti ja toiko sen käyttö mielestäsi hyötyjä kääntämiseen?

5.7. Mitä ongelmia tai vaikeuksia kohtasit kääntäessäsi?

6. Tallentaminen ja puhdistus:
6.1. Oliko projektin tallentaminen helppoa?

6.2. Oliko projektin puhdistus helppoa?

6.3. Mitä ongelmia tai vaikeuksia kohtasit talennuksessa ja puhdistuksessa?

7. Online help ja internet:
7.1. Miten usein käytit online help -osiota?

7.2. Käytitkö internetiä ongelmanratkaisuun?

7.3. Mikä oli strategiasi: a) ensin help-osion tutkiminen ja sitten kääntäminen, b) kääntäminen ja help-

osion tutkiminen samanaikaisesti, c) kääntäminen ja vasta sitten help-osion tutkiminen, vai d) jokin muu, mikä? Vaikuttiko tämä suoritukseesi?

7.4. Löysitkö etsimäsi tiedon helposti ja oliko se selkeää ja helposti ymmärrettävää?

7.5 Mitä ongelmia tai vaikeuksia kohtasit käyttäessäsi online help -osiota?

8. Yleisiä kysymyksiä järjestelmän käytettävyydestä
8.1. Mitkä ovat mielestäsi käytettävyyden suhteen järjestelmän parhaita puolia käyttäjälle?

8.2. Mitkä ovat mielestäsi käytettävyyden suhteen järjestelmän huonoimpia puolia käyttäjälle?

8.3. Olivatko jotkin järjestelmän osat sinusta hämmentäviä tai hankalia ymmärtää?

8.4. Oliko järjestelmässä osia, jotka sinusta olivat ärsyttäviä vaikka ne eivät aiheuttaneetkaan sinulle suuria ongelmia?

8.4. a) Toimiiko ohjelma mielestäsi loogisesti?
b) Aiheuttiko jokin ohjelman toiminnassa sekaannusta tai vaikeuksia?

8.5. Mitkä olivat yleisimpiä virheitä, jotka teit järjestelmää käyttäessäsi?

8.6. Millaisia muutoksia tekisit järjestelmään, jotta se olisi helpompi käyttää käyttäjän näkökulmasta?

8.7. Haluatko mainita mitään muuta?

9. Käännösteknologia kääntäjänkoulutuksessa
9.1. Sisältyykö tutkintoosi käännöstyökalujen käytön opetusta? Jos sisältyy, minkä verran?

9.2. Oletko mielestäsi saanut tarpeeksi opastusta käännöstyökalujen käytössä?

9.3. Oletko sitä mieltä, että koulutus on valmistanut sinut hyvin työelämän tarpeisiin käännösteknologian osalta?

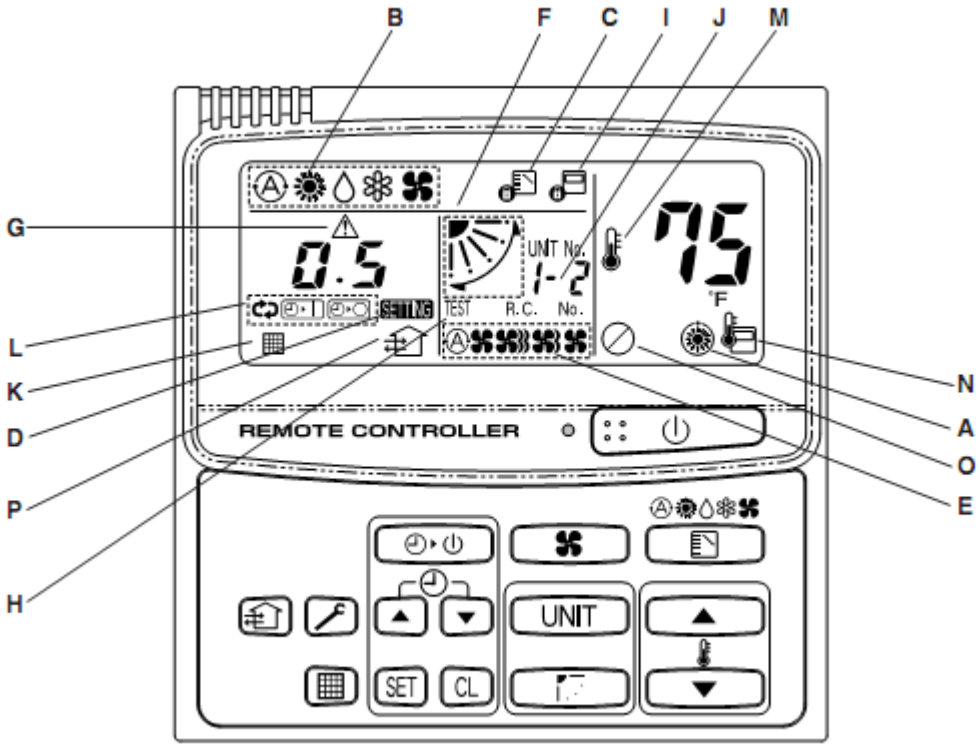
9.4. Tarvitaanko koulutukseen mielestäsi enemmän tai vähemmän käännöstyökalujen käyttöä?

9.5. Muita kommentteja koulutukseen liittyen?

APPENDIX 3. SAMPLE TRANSLATION.

Display (Remote Control Unit)

When you first turn on the ON/OFF operation button, SETTING will blink on the display of the remote controller. During this blinking, the remote controller is verifying units connected. Use the remote controller, after stops blinking.

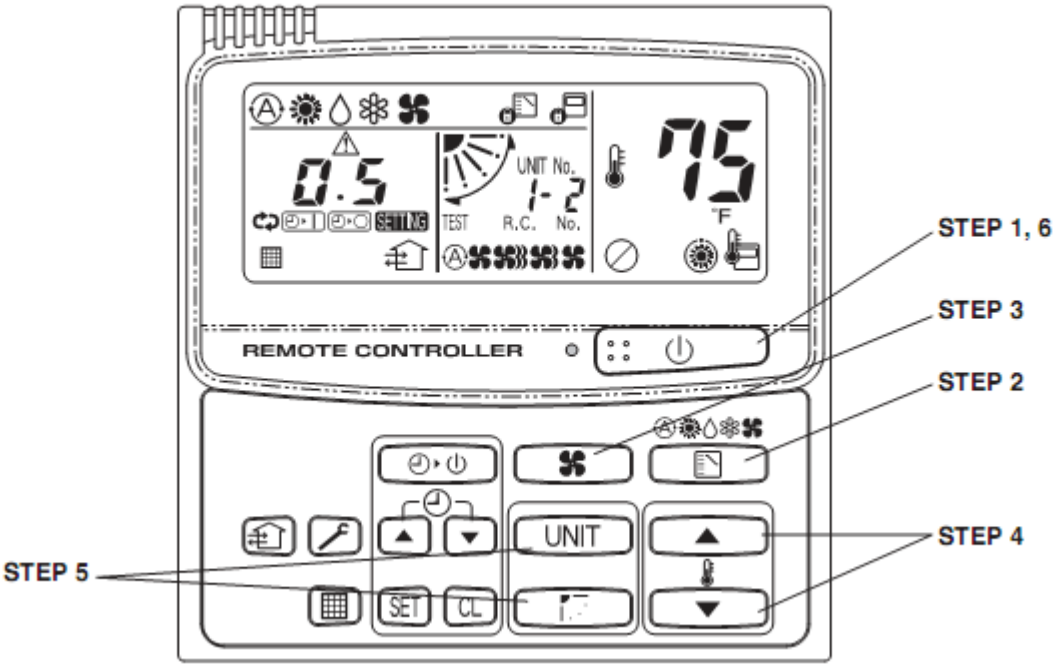


Description

A: When the unit is in the heating standby mode, the * indicator appears. (Refer to the description of the Special Remarks on page 12.)
B: The currently selected operation mode is displayed.
C: This is displayed when a different operation mode was already selected by another remote control unit and
indicates that the mode cannot be changed.
D: When the TIMER SET button is pressed to set the timer, the SETTING indicator flashes.
E: The currently selected FAN SPEED is displayed.
F: The currently selected Airflow Direction and SWEEP status are displayed.
G: This is displayed only when an abnormality occurs within a unit.
H: When the CHECK button is pressed, the TEST indicator appears.
I: This is displayed to indicate that the system controller is being used for control.
When * is flashing on the display, the operation is not accepted by the system controller.
J: This indicates the indoor unit address in case of group control.
Unit No.
1 — 2
↑Indoor unit No.
↑ Refrigerant circuit No.
K: This is displayed when it is time to clean the filter.
L: When setting the timer, the selected timer mode is displayed.

M: This indication appears when the temperature is set.
N: This indication appears when the remote control sensor is used.
O: This indication appears when the ventilation fan does not operate even if the ventilation button has been
pressed.
P: This indication appears when the ventilation fan is operating.

Operation



NOTE: To warm up the system, the power mains must be turned on at least five (5) hours before operation.

STEP 1 **To start the air conditioner**
Press the operation button (ON/OFF button).

STEP 2 **Setting the mode**
Press the MODE button to select the mode of your choice.
[* (AUTO), * (HEAT), * (DRY), * (COOL) or * (FAN)]
NOTE: * (AUTO) and * (HEAT) are used only for heat pump type.

STEP 3 **Setting the fan speed**
Press the FAN SPEED selector button to select the fan speed of your choice.
[* (AUTO), * (HI.), * (MED.) or * (LO.)]
If AUTO is selected, the fan speed switches automatically.

STEP 4 **Setting the temperature**
Use the < or > button as appropriate to change the temperature setting as desired.
(< reduces the temperature, and > increases the temperature.)

STEP 5 **Setting the airflow direction**
When more than one indoor unit is connected, the UNIT button is used first to select a unit. Then use the FLAP button to set the airflow direction to a specific angle or to

sweep. (Refer to the description of the remote control unit.)

STEP 6 **To stop the air conditioner**

Press the operation button (ON/OFF button) again.

APPENDIX 4. TEST INSTRUCTIONS.

Käytettävyystesti: ohje

Tehtävät:

- Avaa WordFast
- Luo uusi projekti ja käännösmuisti, tallenna kansioon Testi (työpöydällä). Nimeä sekä projekti että käännösmuisti samalla tavalla niin, että ne tunnistaa.
- Aseta Fuzzy Match Threshold 65:een.
- Käännä teksti käyttäen luomaasi uutta käännösmuistia. Käännettävä teksti on nimeltään ”testiasiakirja”, ja se löytyy työpöydän Testi-kansiosta. Myös alkuperäinen teksti kokonaisuudessaan löytyy samasta kansiosta, käännettävä tekstinpätkä on alkuperäisen asiakirjan sivuilla 6-7.
- Lopuksi puhdista ja tallenna projekti.
- Vastaa lomakkeessa esitettyihin kysymyksiin. Kirjaa lomakkeeseen kaikki tehtävän aikana esiin nousseet ajatukset, ongelmat ja muut asiat.

Huom: alkuperäisessä asiakirjassa olevat symbolit on korvattu testiasiakirjassa tähdellä (*).

Kääntäessäsi voit käyttää apuna WordFastin omaa Online Help-osiota sekä internetiä.
Muista, että tarkoitus on testata ohjelmaa, ei käyttäjää. :)

Kiitos!